UNIVERSIDADE DA BEIRA INTERIOR
Engenharia

Automatization of Real Time GNSS Data Acquisition, Validation, Processing for Water Vapor Estimation
NUVEM

Hugo Valentim Figueira Gonçalves

Dissertação para obtenção do Grau de Mestre em
Engenharia Informática
(2º ciclo de estudos)

Orientador: Prof. Doutor Rui Manuel da Silva Fernandes

Covilhã, Junho de 2016

Acknowledgements

Java is easier to write than English, but great achievements are only possible with hard work...

Abstract

In the recent years, Global Navigation Satellite Systems (GNSS), in particular Global Positioning Systems (GPS), have proved their capacity to monitor atmospheric water vapor with an accuracy that is comparable to the accuracy of other conventional meteorological sensors. GNSS has several significant advantages compared to the traditional observational systems, including low operating expense, all-weather operability and high temporal/spatial coverage. As a result, the provision of Zenith Total Delay (ZTD) and derived Precipitable Water Vapor (PWV) from GNSS data has become established as a standard and valid technique. These atmospheric parameters are currently assimilated into regional weather forecast models (Meteo Office UK; Meteo France and German Research Center for Geosciences (GFZ)). However, the accuracy of the near real-time estimates for these values is still in active research issue.

This dissertation describes the implementation and testing of the estimation of PWV in near-real time that was carried out at SEGAL in framework of the NUVEM project during 2015-2016. The scheme to estimate the atmospheric parameters (ZTD and PWV) and to deliver them timely to the Portuguese Meteo Office (IPMA - Instituto Português do Mar e da Atmosfera) is described in detail. The focus was on the computational component of the scheme but some studies using the estimated solutions were also necessary to be performed in order to obtain some necessary parameters for the operational phase of the project.

The timely estimation of the solutions depends of several processes, which needed to be analysed and optimized separately (but considering always their integration in the full schema), namely:

- download of the GNSS data from external providers since the project relies on the data acquired by different GNSS networks in Portugal and Spain.

- download of the additional products (orbits and clocks) necessary to process the GNSS data.

- retrieve from numerical models the values (temperature & pressure) necessary to convert from ZTD to PWV (done at IPMA).

- processing of the GNSS solutions as soon as the data and products were available.

- uploading of numerical and graphical solutions to IPMA and to the dedicated website.

- redundancy of the system in order to guarantee a maximum operational (online) time.

Additional features and improvements can be implemented in the future in order to further optimize the use of GNSS-PWV in Portugal for nowcasting. Nevertheless, the goals were achieved since the NUVEM system is currently running in operational mode at SEGAL providing IPMA with timely PWV estimates that can be used to help the analysis of atmospheric events in near-real time.

Keywords

GNSS, Zenith Total Delay, Precipitable Water Vapor

Resumo alargado

Nos últimos anos, Sistemas de Navegação Global por Satélite (GNSS), em particular o Sistema de Posicionamento Global (GPS), têm provado a sua capacidade para monitorizar o vapor de agua existente na atmosfera, com uma precisão ao nível de outras soluções mais tradicionais. Este tipo de sistemas tem várias vantagens comparado com os tradicionais sistemas de observação. Entre estas vantagens estão o baixo custo de operacionalização, e a alta cobertura espacial e temporal. Como resultado a previsão de Atraso Total de Zenith (ZTD) e a previsão de Vapor de Água Precipitável (PWV) foi estabelecido como uma técnica de observação standard. ZTDs e PWVs já são hoje em dia utilizados em vários países (nomeadamente Reino Unido, França e Alemanha) como um dos factores que ajuda nos modelos de meteorologia. No entanto, a precisão das estimativas de ZTD/PWV em tempo quase real baseadas em orbitas ultra rápidas do sistema GPS, encontra-se ainda em estudo.

Nesta dissertação é descrita a implementação e os testes de validação da computação de PWV em tempo quase real feita pelo SEGAL no âmbito do projecto NUVEM que ocorreu entre 2015 e 2016, o esquema de cálculo dos parâmetros atmosféricos (ZTD e PWV) e o seu envio atempadamente ao Instituto Português do Mar e da Atmosfera (IPMA) é descrito em detalhe. O foco centrou-se na componente informática. No entanto alguns estudos, relativos aos resultados obtidos, foram realizados de forma a poder definir alguns parâmetros necessários na fase operacional do projecto.

A computação das soluções está dependente de vários processos. Houve a necessidade destes serem analisados e optimizados separadamente. Eles são:

- download dos dados GNSS dos provedores externos, uma vez que o projecto depende desses dados, adquiridos em diferentes redes GNSS em Portugal e Espanha;

- download dos dados adicionais (órbitas e relógios) necessários à computação dos dados GNSS;

- obtenção via modelos de previsão numérica dos valores de pressão e temperatura necessários para a conversão de ZTD em PWV (feito pelo IPMA);

- computação das soluções GNSS tão rápido quanto possível;

- upload das soluções (numéricas e gráficas) para o IPMA e para o website;

- redundância dos recursos computacionais.

Existem funcionalidades e melhorias que podem ser implementadas no futuro de forma a optimizar o uso de GNSS-PWV para o nowcasting em Portugal. Contudo, os objectivos foram atingidos, uma vez que, o sistema NUVEM está em modo operacional, sendo executado no SEGAL e providenciando ao IPMA as soluções que são depois usados na análise de eventos atmosféricos em tempo quase real.

Palavras-Chave

GNSS, Atraso Total de Zenith, Vapor de Água Precipitável

Contents

List of Figures

List of Tables

List of Acronyms

UBI	Universidade da Beira Interior
FCT	Fundação Ciência e Tecnologia
SEGAL	Space and Earth Geodetic Analysis Laboratory
NUVEM	New methods to Use GNSS Vapor Estimates for Meteorology of Portugal
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
ZTD	Zenith Total Delay
PWV	Precipitable Water Vapor
IPMA	Instituto Português do Mar e da Atmosfera
GFZ	German Research Center for Geosciences
GLONASS	Russia Global Navigation Satellite System
Galileo	Europe global navigation satellite system
BDS	China BuiDou Navigation Satellite System
TOA	Time Of Arrival
ZHD	Zenith Hydrostatic Delay
ZWD	Zenith Wet Delay
MF	Mapping Function
NWP	Numerical Weather Prediction
VMF1	Vienna Mapping function 1
DD	Double Difference
PPP	Precise point positioning
JPL	Jet Propulsion Laboratory
IGS	International GNSS Service
COST	Co-operation in the Field of Scientific and Technical Research
EUMETNET	Network of European Meteorological Services
E-GVAP	EIG GNSS water vapour program
GEO	Geostationary Earth Orbit
LEO	Low Earth Orbit
TPW	Total Precipitable Water
NAS	Network Attached Storage Disks
UPS	Uninterruptible Power Supply
SAN	Storage Area Network
OVMS	Oracle Virtual Machine Server
OVMM	Oracle Virtual Machine Manager
NFS	Network File System
VM	Virtual Machines
OS	Operating System
RAID	Redundant Array of Inexpensive Drives
IGN	Instituto Geográfico Nacional
UTC	Universal Time Coordinated
FTP	File Transfer Protocol
NTRIP	Network Transport of RTCM via Internet Protocol
BNC	BKG NTRIP Client
ReNEP	Rede Nacional de Estações Permanentes
SERVIR	Sistema de Estações de GNSS de Referência Virtuais

Chapter 1

Introduction

1.1 Motivation

Weather forecast is a very difficult field of study, because many factors can alter the weather in a question of minutes. This is clearly a big problem since an extreme weather event can occur without a warning of the responsible authorities. Without a proper warning from the authorities people won't take precaution. This may cause extreme loss, or even, in some very extreme cases, death. Figure 1.1 shows an example of an extreme weather event that took place in Lisbon in 22 September of 2014.

Figure 1.1: Lisbon, 22nd September 2014, extreme weather event

In some countries, particularly in Europe, a new method to help in the prediction of these extreme weather events has been emerging. This method consists in using the Global Navigation Satellite System (GNSS) networks. GNSS stations provide observations at a really fast and continuous rate. With the GNSS observations it is possible to calculate the Zenith Total Delay (ZTD), that represents the atmospheric delay of the GPS signal arriving along the zenith direction. From this it is possible to estimate the Precipitable Water Vapor (PWV) (a more detailed explanation regarding the co-relation between ZTD and PWV is located in Chapter 2). The amount of Water Vapor in the atmosphere is one of the most important factors influencing the weather. This makes the GNSS-PWV product a great help to Meteorologists, so that they can identify extreme weather events more easily.

This dissertation is part of the NUVEM project (New methods to Use GNSS Vapor Estimates for Meteorology of Portugal). This project was an exploratory project financed by the FCT (Portuguese Science Foundation) with the reference, EXPL/GEO-MET/0413/2013, that started in 01 of April of 2014 and ended in 30 of September of 2015.

On the personal side, what motivated me to choose this topic has my dissertation was the possibility of working in a multi-disciplinary project that combine Meteorology and Computer Sciences. This is very important to me because it is a real world application of Informatics that will connect Geodesy and Meteorology. And it is very gratifying being a part of such an important project.

1.2 Problem Statement

The problem addressed in this dissertation concerns the issue partially discussed above. The weather can be very unpredictable, and there are few methods that can provide such an important factor, as Water Vapor in a continuous near real time approach.

In this scenario the GNSS-PWV can be of good help to the responsible authorities. But before they can use this product it is very important to study how reliable it is. Therefore it is necessary to develop an automated system that can calculate the GNSS-PWV and after that validate the system.

1.3 Objectives

The main objective of this dissertation was to optimize the operationalization of the computation of PWV estimates using data retrieved from several GNSS networks in Portugal and Spain and deliver them timely to the Portuguese Meteo Office (IPMA). This process should be done as fast as possible, since the main purpose is to support the nowcasting activities, particularly during the occurrence of Extreme Weather events.

It was also necessary to study the reliability of the networks, the influence of lack of data for the PWV computation and, most importantly, the accuracy of the obtained GNSS-PWV, compared with the most accurate estimates of Water Vapor, that can be obtained about 15 days later.

1.4 Approach

The adopted approach for solving the problem mentioned in the previous section can be summarised and structured into the following stages:

1. The first stage consisted in identifying several elements necessary for the GNSS-PWV computation. Defining strategies to obtain them efficiently, whilst taking into account the SEGAL and IPMA infrastructures;

2. Stage two consisted in preparing SEGAL's infrastructure in order to provide the best support to the project;

3. Stage three refers to the development of the necessary software to automate the system;

4. In Stage four a series of tests were conducted to asses, the quality of the generated solutions, and the reliability of the networks used;

5. Stage five, the last one, consisted in applying the rules, generated based on the tests conducted in Stage four.

1.5 Main Contribuitions

The main contributions of this dissertation are presented here .

The first contribution consists in the development of several computational applications that, based on existing software packages, implemented an integrated system to:

- Download GNSS raw data from several sources;

- Processing the acquired data using the GIPSY/OASIS software (more details further);

- Extract ZTD values;

- Calculate the PWV;

- Upload files to IPMA.

During the course of this dissertation the following posters and conference papers were presented:

- A Poster with the title "New methods to Use GNSS Vapor Estimates for Meteorology of Portugal (NUVEM)" was presented at GNSS4WEC 2014 Summer School, in Varna, Bulgaria in September 2014 (http://nuvem.di.ubi.pt/publicacoes/valentim_varna.pdf).

- A Poster with the title "New methods to Use GNSS Vapor Estimates for Meteorology of Portugal" was submitted and accepted for the AGU 2014 in San Francisco, United States in December 2014 (http://nuvem.di.ubi.pt/publicacoes/fernandes_agu.pdf).

- An article with the title "Aplicações de produtos de vapor de água com base em GNSS" was submitted and accepted for the APMG2015 – 9º Simpósio de Meteorologia e Geofísica da APMG, in Tavira, Portugal in March 2015 (http://nuvem.di.ubi.pt/publicacoes/machado _apmg.pdf).

- An article with the title "Desenvolvimento de um produto de vapor de água integrado com base em estimativas GNSS" was submitted and accepted for the APMG2015 – 9º Simpósio de Meteorologia e Geofísica da APMG, in Tavira, Portugal in March 2015 (http://nuvem.di.ubi. pt/publicacoes /rodrigues_apmg.pdf).

- An article with the title "Operational Forecast in Near Real Time of Perceptible Water Vapor from GNSS data for Portugal" was submitted and accepted for the 9th Spanish Portuguese Assembly of Geodesy and Geophysics, in Madrid, in 28-30th June 2016

A project website with the general info and publications was also developed, and can be found at http://nuvem.di.ubi.pt

1.6   Dissertation Structure

This dissertation is organised in seven chapters. The contents of such chapters can be summarised as follows:

Chapter 1 - Introduction Refers to this chapter, it contains the motivation. The problem statement, the objectives of the dissertation and the approach used throughout the project.. It also contains the main contributions for this dissertation.

Chapter 2 - GNSS and Water Vapor In this chapter the relation between the GNSS and the water vapor is described in detail, namely the definitions and methods of estimating water vapor.

Chapter 3 - State of the Art Analyses the different GNSS orbit types, the different methods to compute GNSS data, and the different software available for it. An explication regarding the different satellite Orbits and Clock products is made. It also analyses different Pressure & Temperature models. A small comparison between GNSS-PWV and other methods to obtain PWV is made. It ends with the methods used to compute GNSS-ZTD in the MetOffice UK..

Chapter 4 - System and Network Infrastructure Preparation In this chapter the preparation of the Server that olds the product, the internal network changes made, and the virtualization infrastructure implemented for a redundant system are described.

Chapter 5 - Near Real Time Data Acquisition The several GNSS data networks and providers is described. It is detailed and explained the system architecture. It is also explained how the different parts needed to generate the GNSS-PWV are obtained.

Chapter 6 - Data Validation In this chapter an analyses on the networks reliability is made. The results of several tests made regarding the missing data in a given time window is analysed, as well as the results of the quality testing for this product. Based on the results obtained, some rules are also defined.

Chapter 7 - Conclusions Presents the main conclusions of this dissertation as well as some directions regarding future work related with this work.

Chapter 2

GNSS and Water Vapor

2.1 GNSS Overview

Global Navigation Satellite Systems (GNSS) is the term that refers to all space-based navigation systems. Currently, there are several systems, namely, the Global Positioning System (GPS) from the United States, Russia Global Navigation Satellite System (GLONASS), Europe's Global Navigation Satellite System (Galileo) and China's BuiDou Navigation Satellite System (BDS). These systems have revolutionised the surveying and mapping industry. This technology involves satellites, ground reference station infrastructure and user equipment. They can estimate the positions and velocity continually and accurately around the world.[RHH05]

Of the several different systems, GPS is the most known and used. With that in mind, GPS will be the primary focus in this dissertation. The GPS satellite navigation system consists in 24 satellites arranged in 6 orbital planes where each plane contains 4 satellites. Like was mentioned before, GPS provides to the normal users positioning and velocity information. It was also designed to be able to transfer time between ground stations. The GPS services can be used by an unlimited number of users because GPS receivers only receive data[Kap96].

GPS satellites broadcast navigation data and ranging codes on four different frequencies: L1 (1575.42 MHz), L2/L2C (1227.6 MHz) and L5 (1176 MHz)[gps11]. Each satellite transmit a different ranging code. The different ranging codes allows the use of a technique called One-way Time of Arrival (TOA). TOA consists in determining the propagation time of the signal and consequently calculate the satellite-to-user range while the navigation data allows each receiver to know the location of each satellite at the transmission signal[Kap96].

It is possible to determine the receiver's three-dimensional location by combining the satellite's location and the TOA ranging measurements. Due to the uncertainty of the time value, four satellites are needed to obtain a position (unless that the receiver is connected to an atomic clock) [Kap96].

2.2 Water Vapor Overview

Water vapor represents a very important role in atmospheric processes that act over a wide range of temporal and spatial scales from climate to micrometeorology[BBH$^+$92]. Several atmospheric processes are influenced by the water vapor distribution (e.g., radiation transfer, energy balance, cloud formation/composition and the convective initiation and development of precipitation systems)[BAD05]. It is also the major constituent in the greenhouse effect. Water vapor is the most variable constituent of the atmosphere and is an essential component of climate and numerical weather prediction because of its relation with the atmosphere

processes[Gue03].

With such an important role scientists aimed to create a series of methods to measure the distribution of water vapor in the atmosphere (i.e., radiosondes, microwave radiometers, spectrometers, lidar and GNSS which is the method described in this dissertation).

2.3   GNSS Water Vapor Estimation

There are several advantages of GNSS water vapor estimation over other existing methods. The most important ones are, good temporal resolution (down to 5 minutes) and both the fact that can be run unattended and work in all weather conditions[BAD05]. Another important fact is that the number of GNSS stations continues to grow every day, around the world.

GNSS signal suffers several propagation effects, that are referenced in the literature as a signal delay, when passing through the atmosphere.The magnitude of those effects depends on the satellite-receiver elevation angle and the atmospheric environment.[Gue03]
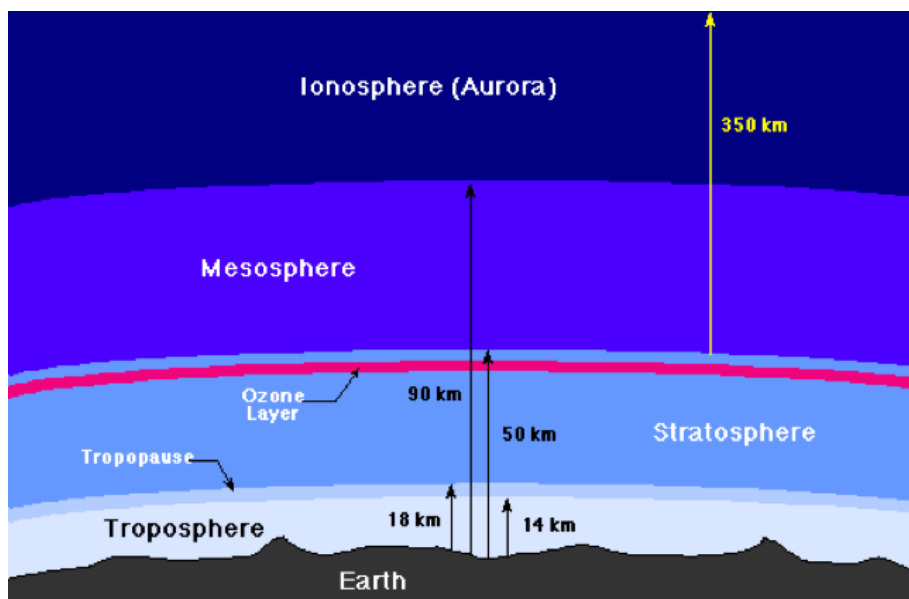
Figure 2.1: Layers of the Earth's atmosphere

Figure 2.1 shows the different atmospheric layers. The delay is mainly due to the ionosphere and troposphere layers, but while the ionospheric delay can be eliminated with Double Difference receivers (because this layer is frequency dependent), the tropospheric layer is not frequency dependent (above 15Mhz), so it can not be eliminated[Gue03].

The Zenith Total Delay (ZTD) is defined as the atmospheric delay of a GNSS signal that would arrive from the zenith direction.

Figure 2.2 shows an illustration of the GPS signal, zenith direction and elevation angle (e). ZTD is the sum of two delay effects in the troposphere. Those effects are the Zenith Hydrostatic

6
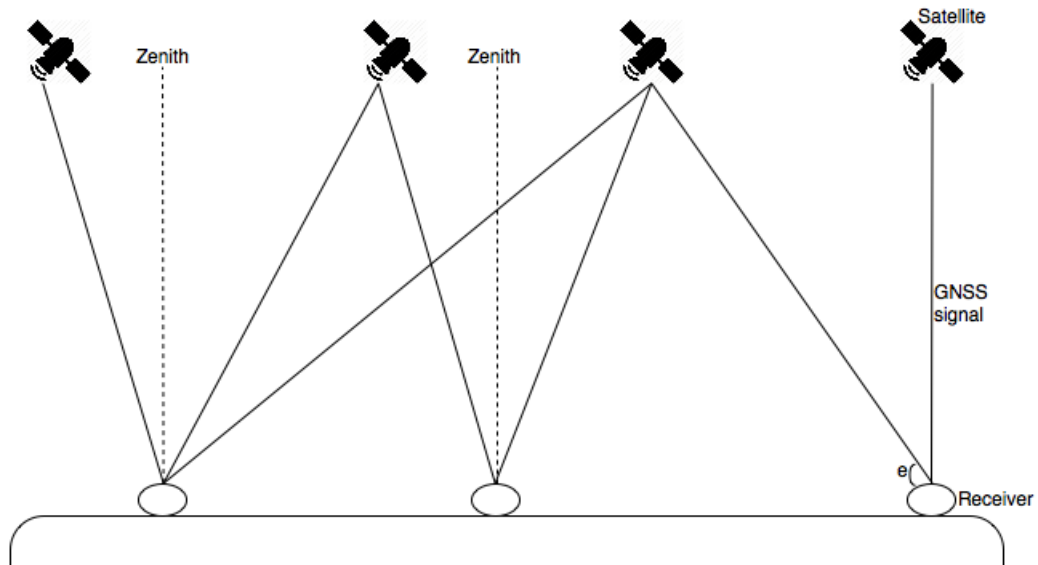
Figure 2.2: GPS signal between the satellites and receiver.

Delay (ZHD) and the Zenith Wet Delay (ZWD). The ZHD is the one causing the bigger impact, and it is caused by dry air containing mainly N2 (nitrogen gas) and O2 (oxygen). The ZHD represents a delay of about 2.1 meters at sea level and it varies mainly with the temperature and atmospheric pressure. This fact makes it predictable. The ZWD oscillate normally between 1 and 80 centimeters. However, ZWD has a higher spatial and temporal variation than the ZHD making it more difficult to predict[Gue03].

In order to model the tropospheric delay precise GNSS positioning requires an accurate Mapping Function (MF). There are several mapping functions. Along the years the most used were the Niell mapping function[Nie96], the Global Mapping Function[BNTS06], and this last one successor GPT2[LSB+13]. There is however a problem with these functions. They all are based in climatology. That fact makes them have a limited capacity to predict short-term and anomalistic atmosphere behaviour.The Numerical Weather Prediction (NWP) models capture the short-term variability of the atmosphere. This fact makes the NWP be considered as a good data source from mapping functions[Nie01].

At the time of the writing of this dissertation the Vienna Mapping Function 1 (VMF1) is considered as the most accurate available[TBHS07]. What might explain its great accuracy is that it is based on an accurate functional formulation and an accurate underlying atmospheric data source, namely the operational analysis from the integrated forecast system of the European Center of Medium range Wheater forecast[BWS06].

The ZHD can be calculated easily and with great accuracy, using measurements of surface atmospheric pressure with the following formula[Saa72] [AN87]:

$$ZHD = 2.2768 \pm 0.0015 \frac{P_0}{f(\phi, H)}$$

$$f(\phi, H) = 1 - 0.00265 cos(2) - 0.000285 H$$

where P_0 is the surface pressure in millibars, ϕ is the latitude and H is geodetic height in kilometres. ZHD is given in millimetres.

Calculating ZHD and knowing ZTD (which is the value estimated by the GNSS processing) we can calculate the ZWD using the formula:

$$ZTD = ZHD + ZWD$$

After the ZWD has been calculated it can be multiplied by the variable $\prod$ [AN87] to obtain the quantity of water vapor integrated in the zenith direction (IWV):

$$IWV = \prod.ZWD$$

where $\prod$ is given by:

$$\prod = \frac{10^6 m_w}{(k_2 - k_1 \frac{m_w}{m_d} + \frac{k_3}{T_m})R^*}$$

where T_m is the weighted mean temperature of the atmosphere, m_w is the molar mass of water vapor, m_d is the molar mass of dry air, R^* the ideal gas constant, k_1, k_2 and k_3 are physical constants that are based in part on theory and in part on experimental observations[BBH$^+$92]. According to [BBC$^+$94]:

- $k_1 = 77.60 \pm 0.05 K/hPa$

- $k_2 = 70.4 \pm 2.2 K/hPa$

- $k_3 = (3.739 \pm 0.012).10^5 K^2/hPa$

Bevis et. al in [BBH$^+$92] have derived the following formula:

$$T_m = 0.72t + 70.2(^oK)$$

2.4   Summary

In this chapter an overview to the GNSS and Water Vapor was presented. It was also shown how the GNSS Water Vapor estimation is calculated. This chapter was widely influenced from chapter 5 of [Ben13].

Chapter 3

State of the Art

3.1 GNSS Processing

3.1.1 Processing Methods

There are two methods for processing GNSS raw data, Precise Point Positioning (PPP) and Double Difference (DD), which will be discussed in the following subsections.

3.1.1.1 Double Difference

Double Difference, differently from PPP, is a network processing approach. The idea is to form baselines between GNSS stations, with this method it is possible to eliminate the clock errors. Double Difference results of the combination of two other methods:

Between-Receivers Single Difference: In this technique there are two GNSS receivers observing the same satellite at the exact same time. This method provides better positions estimates by subtracting each receiver observations equation from the other because common error sources are cancelled out when the difference is taken.[SD14]

Between-Satellites Single Difference: In this case the idea is to have just one receiver observing two satellites. What is subtracted here is the code and/or phase measurement of one satellite from the other. The data obtained in this method eliminates the receiver clock error (since there is only one). This method does not provide a better position estimate.[SD14]

With the combination of both methods the benefits of both are acquired, the improvement of position estimates gained through the Between-Receivers Single Difference method and the clock error suppression obtained from the Between-Satellites Single Difference method. Not just the receiver clock error is eliminated, but also the satellite.[SD14]

3.1.1.2 Precise Point Positioning

PPP is a method for performing precise position determination using a single GNSS receiver. This positioning technique is based on the fact that precise GNSS orbit and clock data products, with centimeter accuracy, have been widely disseminated in recent years. That data can be used to reduce drastically the errors in the GNSS satellite orbits and clocks. Those are two of the biggest error sources in GNSS positioning.[Gao06]

If a dual-frequency GNSS receiver is combined with satellite clocks and precise positions, PPP will be able to provide solutions with an accuracy of centimeter. That fact might be appealing to several applications. Unlike double-difference positioning that requires observations from at

least one base station with known coordinates and will be discussed in the next subsection, PPP does not require such observations from a base station. PPP is distinguished from the conventional point positioning techniques, because of the precise part. The other ones only use code or phase-smoothed code as the principal observable for position determination.[Gao06]

There are a few advantages of choosing the PPP technique over the differential precise positioning techniques:[Gao06]

1. The fact that PPP only requires one GNSS receiver means that establishing a base station is not necessary. This eliminates the constraint of simultaneous observations as well as the operating range limit, on both rover and base receivers imposed by the differential RTK technique.

2. Other great benefit, when compared to the differential approach, is the reduced equipment cost, and labor, because the use of base station(s) is not necessary.

3. PPP can also be applied in other fields, since PPP needs to estimate receiver clock and tropospheric effect parameters, it provides a new way for precise time transfer and water vapor estimation using only one GNSS receiver.

A downside about this approach is that if there is an error in the clock of the GPS satellite, that error will propagate in a larger position error at every station.

3.1.2   Processing Software

There are several softwares for GNSS data processing. There are different types of licensing and different feature in them. In this section it can be found a slight overview about them.

First of all we take a look at the licensing type of all the softwares:

Software	Type of Software	Licence Details
RTKLIB	Open-Source	BSD 2-clause license
GPSTk	Open-Source	GNU LGPL
BERNESE	Commercial	
RTNet	Commercial	
GAMIT	Commercial with Free Licence for Universities and non-commercial uses	
GIPSY-OASIS	Commercial with Free Licence for Universities and non-commercial uses	

Table 3.1: Types of Software and Licencing

A more detailed analysis on these software options are taken in Appendix A.

3.1.2.1   GIPSY-OASIS

GIPSY-OASIS, is the GNSS-Inferred Positioning System and Orbit Analysis Simulation Software package. GIPSY is developed by the Jet Propulsion Laboratory (JPL), and maintained by the

Near Earth Tracking Applications and Systems groups.

Its main characteristics are[gip]:

Analysis of data from Global Navigation and Satellite Systems: GPS, GLONASS, DORIS, SLR;

Precise centimeter-level GNSS-based positioning and timing: Space platforms, including low-Earth orbiters and GNSS constellations, Airplanes, Terrestrial (ground) stations, static or moving and Time transfer;

Single-receiver ambiguity resolution: using JPL's GPS orbit and clock products;

Terrestrial positioning for geophysical research: Earth deformation, plate tectonics, GNSS Networks, Ice flow, Climate studies through observation of troposphere and ionosphere, Reference frame (geocenter and scale) and Earth rotation parameters;

Airplane positioning

Precise orbit determination: Low-Earth orbiters: Topex/Poseidon, Jason-1, Jason-2/OSTM, GRACE, Champ, SAC-C, COSMIC, Space Shuttle (e.g. SRTM), Commercial Satellites; GNSS constellations including GPS and GLONASS.

3.1.3 Method and Software Used in the Dissertation

With the descriptions above, regarding the different software solutions, it is easy to conclude that all of them would be able to perform the necessary tasks for this specific project. However the learning curve of these software and their basics, are big. It normally takes a few months just to learn the basics in each and every one of them.

So with these problems, it was established, at the start of the project that the software used would be the GIPSY-OASIS. The reason, is that the SEGAL team, and mostly Prof. Rui Fernandes, have been working with GIPSY for over 10 years. This sure helped keeping the project on time.

Since GIPSY-OASIS was used, the Processing Method was the PPP, since it is the only one used by GIPSY.

3.2 Satellite Orbits

One of the most important components for every GNSS based processing are the Orbits of the satellites. There are two providers of orbit observations the Jet Propulsion Lab (JPL) of NASA[jpl], and the International GNSS Service (IGS)[igs].

Both of the providers make Orbit files available in three types:

1. Final - These are the most precise orbit files. They have an accuracy of 2,5cm and are available with a delay of 14 days;

2. Rapid - The Rapid orbits have the second place in terms of precision. They have an accuracy of 3,5cm and are available much sooner than the Final ones, with a delay of just 1 day;

3. Ultra-Rapid - The Ultra-Rapid orbits are a different case. Depending on the provider (JPL or IGS) they contain different estimates. These will be discussed below.

3.2.1 Ultra-Rapid

The Ultra-Rapid orbits are the ones used for this research since they are faster ones (excluding streams) available and we are interested in near real-time water vapor estimation. However, there are two options for these kind of orbits.

The IGS Ultra-Rapid orbits files contain a 48 hours window, the first half (24 hours) is based on observations and the second half are predicted. These files are updated four times a day (03:00, 09:00,15:00 and 21:00 UTC time). When the file is made available they contain six hours in the future (i.e. the file of 03:00 contains orbits (predicted) corresponding to 09:00 of the same day)[igs].

Differently from the IGS Ultra-Rapid orbits, the JPL ones completely based in observations. They contain a 30 hour window with an accuracy of 5cm. Since these are observation they have a delay, in this case of 2 hours and are updated every hour[jpl].

After knowing both possible solutions, it can be observed that the choice was between a more accurate one (the JPL since is based in observations), or a faster one (IGS file has predictions in the future). By now the ideal would be to test with both options and see how much unreliable the IGS predictions would be. However a problem was encountered. GIPSY can use both Final and Rapid orbits form IGS, but it cannot use the Ultra-Rapid ones. What causes this is that GIPSY needs not only the file with the orbits but also a file with the clocks of the satellites. By the time the Rapid and Final orbits are available those clocks are available and so GIPSY can get them and use the IGS orbits. Since the IGS Ultra-Rapid orbits are predicted the same method cannot be applied. So it is here established that for this product the Orbits used are the JPL Ultra-Rapid, and that based on that the solutions generated will have a two hour delay.

3.2.2 Stream

Another possible choice would be to collect the GPS satellite data streams. There are different providers making these available. Some software (like BERNESE) can use these streams input instead of the orbit files and therefore have a fast and reliable(since it is made based on observations) solution for the Real Time products. Unfortunately by the time that this dissertation was written GIPSY (since it only uses the PPP technique) was not capable of using these streams, because the lack of the required Clock corrections.

3.3 Pressure & Temperature

Pressure and Temperature values are necessary variables to convert the ZTD in PWV. The ideal solution, would be that the GNSS stations would be fitted with high precision Temperature and Pressure sensors so that the values used in the conversion would be the most accurate as possible. Since our GNSS data providers did not have their stations fitted with these kind of sensors, and since it would be too expensive to acquire them in the scope of this project (it would cost around 2000€ for each station) it was decided to use Pressure and Temperature predictions from

Numerical Weather Prediction Models.

IPMA, since they already use these models, were responsible for providing them. They suggested two models:

AROME: It is a non hydrostatic numerical prediction model of high resolution in time and regarding a delimited area. It has been developed and maintained by Météo-France. The AROME is executed locally at IPMA with a vertical resolution of 46 levels and a horizontal resolution of 2.5Km[ipm15a].

ECMWF: it is a global model, meaning that it perform a worldwide prediction. The deterministic version is executed with a maximum reach of 10 days. It has a vertical resolution of 137 levels and a horizontal resolution of 16Km[ipm15b].

After a set of preliminary tests conducted at IPMA they decided it would be best to use the ECMWF models. In [KIA+12] is an example case study that also supports that ECMWF provides better results (in most cases) when compared with AROME.

3.4   Water Vapor by other methods

PWV have been being estimated for years, by different methods. In this section it will be taken an overview over the different methods available regarding those methods. There are more methods than the ones presented, but we will take only a look on the ones used at IPMA.

3.4.1   Radiosonde

The idea of the radiosonde is to connect an electronic equipment to a weather balloon, normally of 600gr, filed with helium and launch it. This equipment is equipped with high precision sensors that measure the atmospheric pressure, temperature and the relative humidity of the air and wind in the vertical atmosphere. It is also equipped with a radio transmitter, that sends the data measured by the sensors to the base[ipm15d].

The information obtained can be seen in real time, or stored for studies in the future. Also the data obtained is used to[ipm15d]:

Initialization of the Numerical Weather Prediction Models;

Local prediction of the stability of the atmosphere and the distribution of the Water Vapor;

Studies regarding the atmospheric pollution;

Prediction of the effects of the atmospheric refraction in the propagation of the electromagnetic radiation and of wave sounds;

Studies of the ballistic precision and reach, by calculating the density of the air for a given pressure;

Climate studies.

In Portugal there are three sites where these kind of equipment is launched[ipm15d]:

1. Lisbon;

2. Funchal, Madeira;

3. Lajes, Azores.

Compared with the GNSS solution its main downsides are the spatial coverage, and the fact that they are expensive. In fact only three a day are launched. In Figure 3.1 is an image of a weather balloon carrying a radiosonde.

Figure 3.1: Weather balloon carrying a radiosonde.

3.4.2   Satellite Image

Since the beginning of the firsts satellites launches, it was possible to verify that the observation of the Earth could help, a lot, the Meteorological field. The first satellite completely dedicated to meteorology was launched by the United States in 1960.

Based on its orbit there are two kinds of satellites. The first are the Geostationary Earth Orbit (GEO). They are positioned in the equatorial plan at about 42000Km of height based on the center of the earth. Their orbit is calculated so they can have a velocity equal to the angular velocity of the rotation of the Earth, making so that the satellite always observe the same area. The Meteosat satellites from the EUMETSAT (www.eumetsat.int) are in this category. In the EUMETSAT website it is possible to view, images from the Meteosat9 with a temporal resolution of 30 minutes. Those images are available in two types[ipm15e]:

- Visible Chanel, with a spatial resolution of 1Km (figure 3.2);

- Infrared Chanel, with a spatial resolution of 3Km (figure 3.3).

The second kind of satellites are the Low Earth Orbit (LEO). The orbits for these ones is almost vertical to the poles, and they are locate below 1000Km of height, and provide spatial resolutions of 1.1Km. They make 2 world tours a day, making two passages on the same site every day. Like the GEO satellites they can provide Visible and Infrared images[ipm15e].
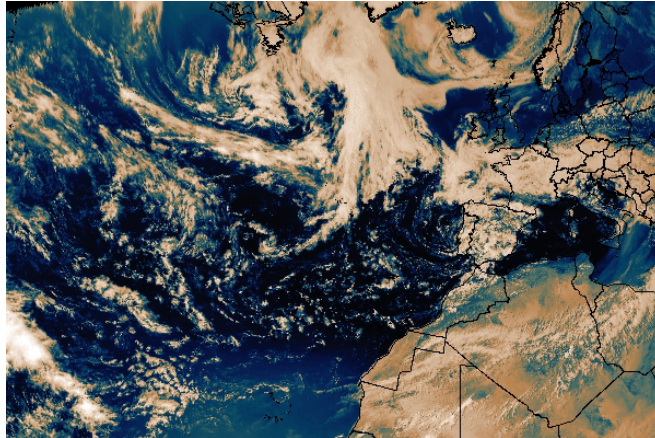
Figure 3.2: Example of a visible Satellite image of the atlantic ocean.
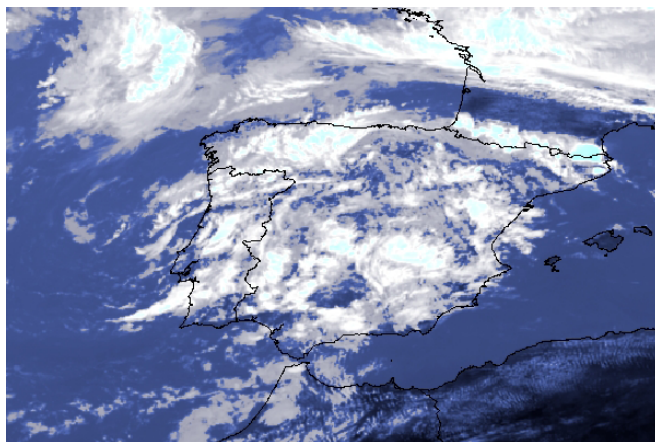
Figure 3.3: Example of an Infrared Satellite image of the Iberian Peninsula.

3.4.3   Numerical Weather Prediction Models

NWP models have a very important role in weather forecasting. NWP is present as a resource in most of the meteorological services in the world. NWP resorts to computer calculations to produce estimates of future states of the atmosphere, using prediction models.

Based on the prediction model, a prediction can be obtained initializing the NWP with some variables (that can be obtained with a radiosonde), temperature, pressure, wind and humidity.

The inherent uncertainties regarding the knowledge of the state of the atmosphere at a certain time, influence the quality of these methods. Those uncertainties and the use of observations irregularly distributed over the globe are some of the error factors that are both introduced in the analysis.

A team started this activity, at IPMA, in 1993. Portugal (by IPMA) is a member of the European Center for Medium-range Weather Forecasts and a partner in the ALADIN project. Currently at IPMA there are three NWP methods used in the weather forecasts, ALADIN, ECMEF and AROME[ipm15c].
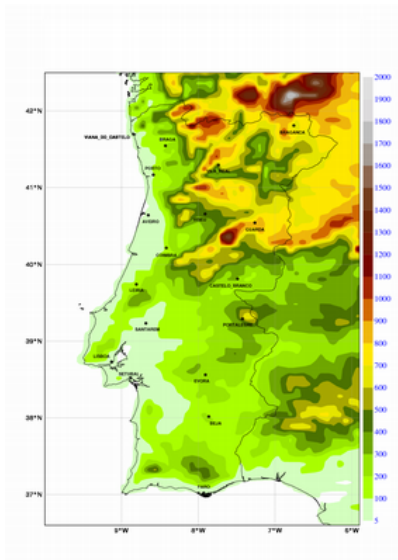
Figure 3.4: Orography based on ALADIN.
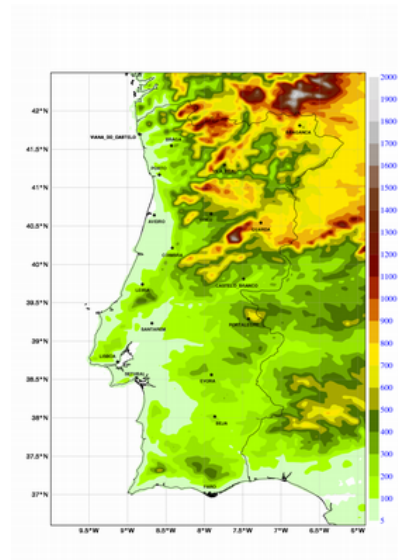
Figure 3.5: Orography based on AROME.
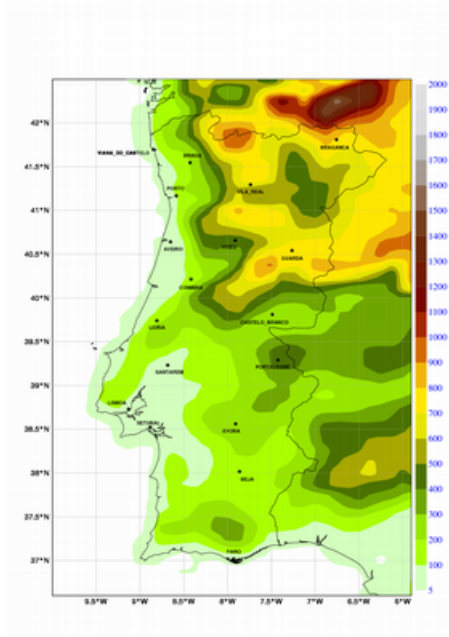
Figure 3.6: Orography based on ECMWF.

3.4.4 NESDIS Blended Total Precipitable Water

Another product at use in IPMA is the NESDIS blended Total Precipitable Water (TPW). This product is derived from multiple sensors and satellites. These products are merged from both, Low Earth Orbit, and Geostationary satellites. The blended methodology provides global coverage with no gaps that allow the analysis of atmospheric moisture over land and ocean. These products have a 16Km resolution over the equator. In terms of products, they have a 16Km resolution, are centered at the equator and 160 degrees West. It has 1437 lines and 2500 elements, which covers the earth from about 71 degrees North to 71 degrees South. The products are available on an hourly basis[nes16].
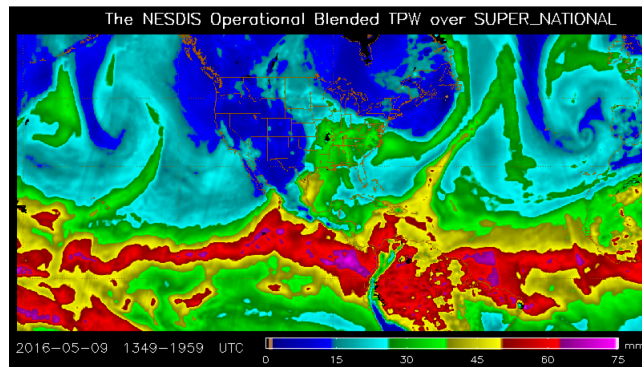
Figure 3.7: Example of an image of the NESDIS project.

3.5 Use Cases around the world

3.5.1 European Project

In 1971 it was created an European intergovernmental framework for cooperation in science and technology. This organization goes by the name Co-operation in the Field of Scientific and Technical Research (COST), and aims for the cooperation between scientific fields, and country cooperation.

In 1998 COST, launched a project (COST Action 716) with the name 'Exploitation of ground-based GPS for Operational Numerical Weather Prediction and Climate Applications'. This project aimed to connect the fields of meteorology and geodesy, in a way to operationalise, and test, the use of ground-based GPS for meteorological applications. The idea was that the GPS stations would provide near real-time observations to be assimilated in the numerical weather prediction. The project ended in 2004 and the result were positive, since the test indicated that ZTD based on ground based GPS had a good accuracy[exp11].

In 2005, a programme called EUMETNET (Network of European Meteorological Services) EIG GNSS water vapour program (E-GVAP) started with the goal of providing its members with ZTD and Water Vapor (WV) estimates in near real-time. How is this achieved? The GNSS data is collected for several of GNSS analysis centers. These centers process that data and forward the obtained ZTDs to E-GVAP dataserver. In this dataserver the ZTDs go through a process of validation, and then are forward to the Meteorological institutes[COS].

3.5.2 METEOOFFICE - United Kingdom

Dr. Jonathan Jones, from Meteo Office - UK, was a member (external consultant) of the the NUVEM project. He provided vital help in some phases of the implementation. In this section a look at the METEOOFFICE - UK ZTD/PWV products will take place. This analysis is based within the testimonial, and PhD. Thesis of Dr. Jonathan Jones.

METEOOFFICE - UK, started in this field in 2002. A partnership with the Institute of Engineering, Surveying and Space Geodesy (IESSG) at the University of Nottingham was made, to develop an automated processing system managed by Met Office staff. This was the first system imple-

mented at the METEOOFFICE - UK, and was given the name of METO. This system was moved from "development" to "production" in 2007. Since the beginning of the METO project tests, regarding the assimilation of the ZTDs in the Numerical Weather Prediction (NWP) models have been performed. Those tests shown that assimilating the ZTDs, in almost all cases have a neutral or positive impact. The reduction of the standard deviation, of the humidity model reached 4% in the several cases.[Jon10]

Since then the METEOOFFICE - UK have implemented three different systems for ZTD assimilation in NWP.

3.5.2.1 METO & METG

The METO and METG system, are basically the same system with just some minor twists. The biggest difference is that the METO concerns to an Europe based network. The METG concerns to a global network. The idea of these systems is to obtain hourly GNSS files from several FTP (File Transfer Protocol) servers and then process these files.

In terms of the processing, there is another difference to the systems. In the METO the system waits 20 minutes (time for it to collect the GNSS files) before starting to process them, in the METG, this delay is of 40 minutes. The METO processing takes around 20 minutes, and the METG takes around 10 minutes (thought it concerns to the whole world it has fewer stations). So the METO ZTDs are ready for assimilation at around HH:40 and the METG ZTDs at HH:50.
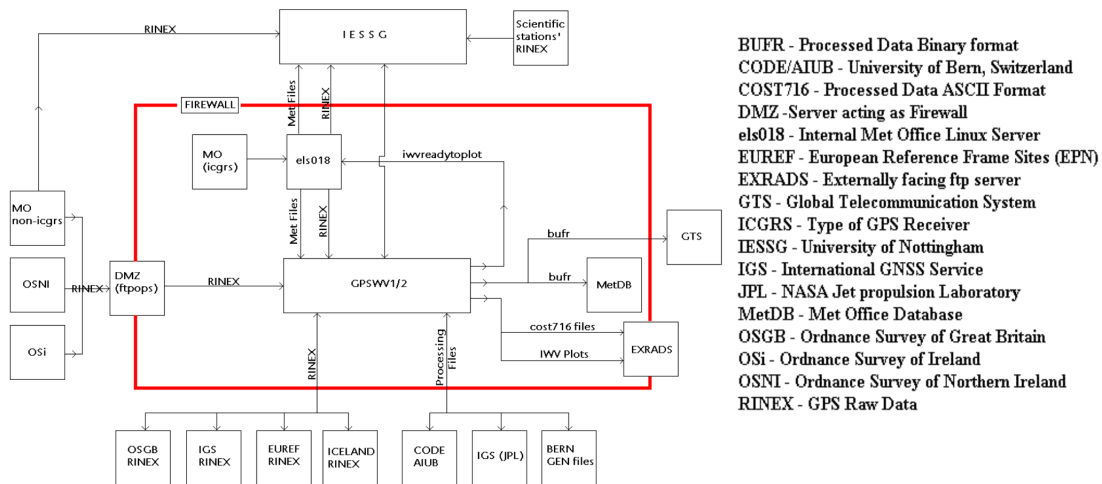
Figure 3.8: METO system processing data flow.

In Figure 3.8 it can be observed the METO system processing data flow. This Figure was taken from the PhD. thesis of Dr. Jonathan Jones.[Jon10]

3.5.2.2 METR

Then there is the METR system, that has a completely different approach. In this system the idea is to collect Real Time data streams from the GNSS stations. This data is streamed via the

NTRIP protocol, and collected with a BNC client. After collecting that data, GNSS files containing 15 minutes are written (4 files per hour). The processing of these files start after 2 minutes (of the 15 minute interval) and the process takes around another 2 minutes. So the average delay for this product is 19 minutes. It is also worth mentioning that this system only concerns to a UK network. And all three systems are process using the Double Difference technique and the BERNESE software.

3.6 Summary

During this chapter an overview on the several components necessary to calculate GNSS-PWV was made. It was also made a description on other methods to obtain Water Vapor estimations used at IPMA. Finally an analyses on the use of both Europe (through E-GVAP) and in the United Kingdom, of these kind of products was conducted.

Chapter 4

System and Network Infrastructure Preparation

A server was acquired at the beginning of the NUVEM project to be fully dedicated to the NUVEM tasks. This server (called meteo) was installed as part of the SEGAL computational infrastructure. After the development of the project, this server will be responsible for the operational part of NUVEM.

After the system was made ready, and already some tests were being made, a hardware malfunction stopped it for a few days. At this point it was decided that, once in production, this should not happen again and it was decided to rearrange SEGAL infrastructure. In this section we will take a look at the changes made, and how the new implementation makes the whole system more reliable.

4.1 SEGAL Infrastructure

SEGAL has a set of 6 Physical Servers, 2 Network Attached Storage Disks (NAS), and 2 Uninterruptible Power Supply (UPS). For the project only the Server bought specifically to the NUVEM project, and one of the SEGAL public server(geodac) were necessary.
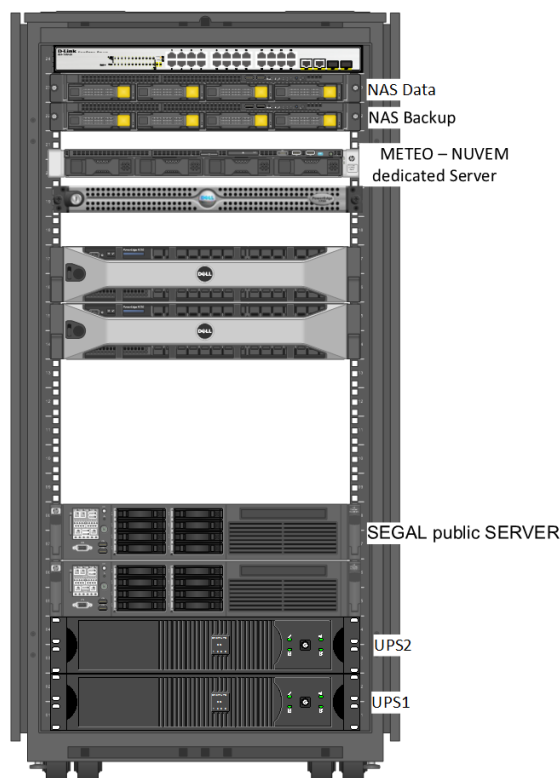
Figure 4.1: Old SEGAL Cabinet and arrangement

Figure 4.1 shows the old disposition of SEGAL's servers (only the elements marked by name where important to this project).

4.2 Virtualization

Two possible solutions were considered after the analysis of the problem. The first one consisted in having a new server and have a duplicated system at all times. This would imply to buy an additional new server (preferably an exact one like the one already in use). Apart from the cost of the new server, the biggest problem, was the waste of resources, since this clone server would be in idle state unless the first one had some kind of malfunction. Besides the fact that it is difficult to have a complete clone all the time (it always take some times to clone all the new data), it is also costly in terms of bandwidth consumption.

The second option was to implement for a virtualization infrastructure. With the resources already available at SEGAL, it would be possible to do it without extra cost. In addition, the virtualization it would make it possible to take maximum advantages of the servers hardware since one server can support multiple Virtual Machines. So at the moment, in terms of costs this seemed the better solution.

4.2.1 Virtualization Approach

In terms of virtualization, there are several implementation approaches to choose from. The more known and "simple" implementation of virtualization, is the capability of one physical machine to run several Virtual Machines. This is a great advantage in terms of hardware optimization. However, by itself this would not provide the solution for our problem, in fact it could make it worst, because if one physical server died, several servers would stop functioning instead of one. A possible remedy for this would be to have two virtualization servers and have a clone of the Virtual Machine. This may seem a little bit similar to the approach of having two physical servers, but it could save SEGAL the money on a new server, because it would be possible to migrate several of the physical servers to virtual machines, and then reuse the free hardware to hold the new virtual machines. There would still have the problem of the need for synchronization of the Virtual Machines between the servers.

Although the scenery described above was a better solution than having a cloned physical server, in the last few years, virtualization techniques have been evolving, and there is a new approach that would not just provide the solution for the problem at end, but also optimize the use of the available resources. This solution has several names, depending on the virtualization provider, but the selected provider (Oracle) uses the name of Server Pooling [pol15]. The idea is actually quite simple. One only requires two or more physical servers and a NAS or a Storage Area Network (SAN) so that the Virtual Machines would be stored on this shareable network drives. The system consists on having the Virtual Machines stored in the network storage and then then the pool servers access this storage to run the different Virtual Machines. This solves the problem of the synchronization because if we have two servers (A and B) and our Virtual Machine is running in server A and that server stops working, the Virtual Machine could run on server B, and no data would be lost because the storage is not in the server that just died. Not just that, in

a Server Pool the management of the available resources is made by the system, meaning if Virtual Machine A is running in Server A and that server dies, it would pass to server B without the need of any interference from the SysAdmin.

But there are also some disadvantages. This approach makes the systems a little bit slower (mainly because it is slower to have the storage in the network than having an Hard Disk in the server), and in a more detailed observation of the available resources at SEGAL, it would mean that the storage in the servers used in the Server Pool would stop being used, since that with this approach, the only thing that can be in the internal Hard Drives is the Operating System. Balancing all advantages and disadvantages, server pooling was the chosen approach that was implemented. Figure 4.2 shows an example of implementation of a server pool implementation.
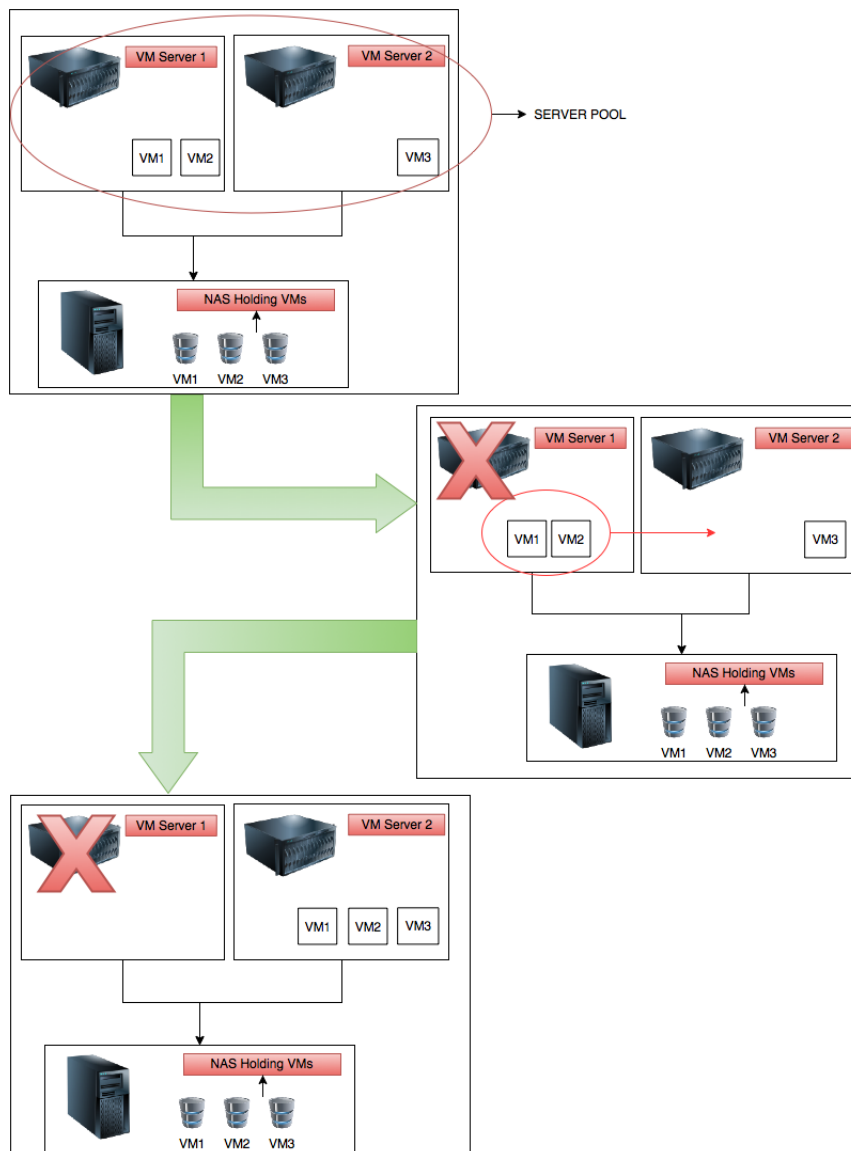
Figure 4.2: Exemplification of the Server Pool System

4.3 Virtualization Hosts

There exist several virtualization software packages. The characteristics of the main ones are discussed in this section.

4.3.1 VMWare

The first provider that was identified was VMWare. VMWare is the market leader in terms of virtualization. A lot of the technologies and features in terms of virtualization were developed by VMWare.

The VMWare has several products, almost all of them with paid licenses licences. The exception is the VMWare ESXi suite. The problem with the VMWare ESXi suite is that it only offers simple features, and not the ones described before. Therefore, and since the cost of the paid products were too high for SEGAL, VMWare was ruled out.

4.3.2 Microsoft Hyper-V

Another provider is Microsoft's Hyper-V which is used by University of Beira Interior Informatics and has all the necessary features. Another advantage of the Microsoft Hyper-V was the existing experience at UBI. The learning curve, and implementation phase, would have been a lot simpler. The problem once again was the licensing, even though it was a lot cheaper than the VMWare solution it was still more than SEGAL, or the project could cover so this option was also ruled out.

4.3.3 Oracle VM

Between several products available, the Oracle Virtual Machine Server (OVMS), was what we needed. It has all the features desired, it is free (paid support) and another advantage it is a Linux System, so it would be simpler, since all the servers in SEGAL are Linux based.

Apart from that, the Oracle VM community is very big and active, and the oracle documentation is very good. This made all the learning on the system really easy. The biggest drawback encountered so far, is that since this is the free version, some bugs take sometime to be solved (in the paid version it only takes a few days, normally, to have these fixes available).

4.4 Oracle VM - Implementation

It was necessary to have the necessary hardware prepared. So, a computer was prepared with Oracle Linux 7 where the Oracle Virtual Machine Manager (OVMM) was installed (any computer with 8Gb of RAM and a x64 bit CPU is eligible for this). A Server (at first we started with only one server, and as long as the migrations started the other ones were added to the server pool)

with the OVMS, and finally creating a Network File System (NFS) share in one of SEGAL's NAS.

Once all the requirements were met was time to follow the procedures, once again, very well documented by Oracle, and create the Server Pool[cre11]. When this was accomplished successfully the migrations could start. As a note, the whole system is then managed via the OVMM (a web interface that can be seen in Figure 4.3).
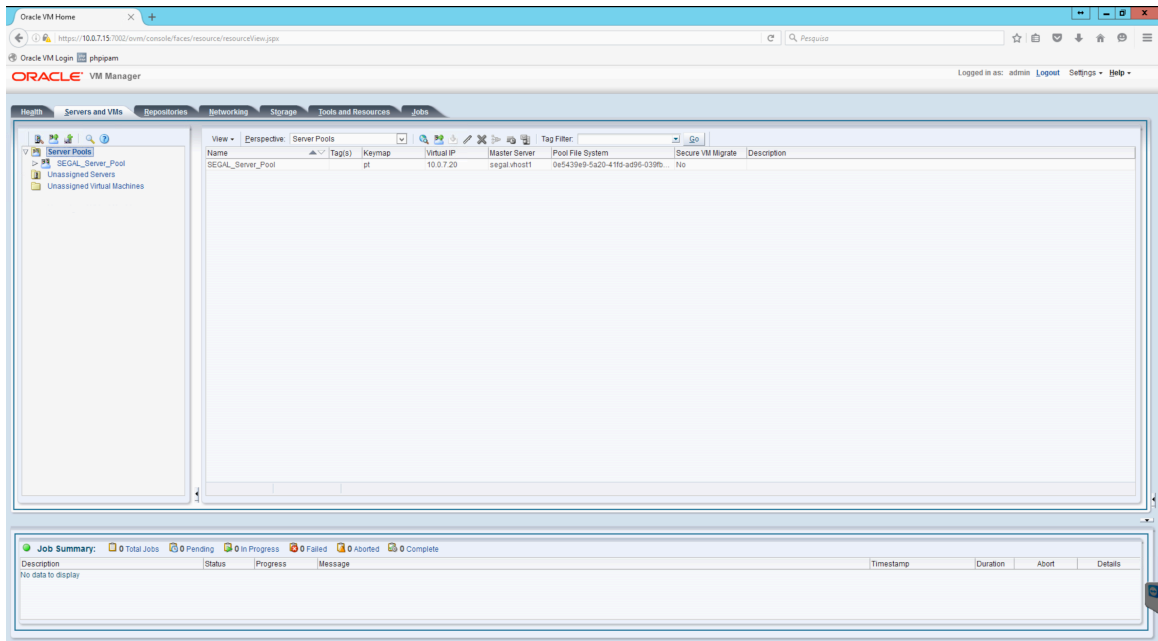
Figure 4.3: Overview of the OVMM web interface

4.4.1 Migration

After the virtualization system started to run, it was time to convert the "old" machines into Virtual Machines (VMs). The first idea was to create clean VMs, and then reinstall and reconfigure that VMs accordingly to the systems they were supposed to replace. Although this might be the best solution, it is not the faster/easier one. Data loss may occur, or even some configurations/services lost during the process, and it would also took a lot more time to reconfigure and reinstall everything.

However, OVMS has a tool dedicated to do the migration. And in a few steps (and some hours later) any "traditional" system can be converted in a VM[con11].

The steps for migrating a physical server to a VM are as follows:

1. Inserting the Oracle VM Server CD in the desired server, and start it by that CD;

2. When the system reaches the boot menu, type "p2v";

3. Then it is given a menu to insert Network details, this should be in the same network as the Oracle VM Server;

4. Then is logical disk(e.g., partitions, apart from the partition that holds the operating system, the user can only choose the desired ones) selection (you can choose all, but if

there is any not necessary you can discard it);

5. After a new menu prompting for the Virtual Machine you want to create for this server (VM Name, Virtual CPUs, Memory and password);

6. Finally go to the OVMM platform, select "import from template" and in the field "Template URLs" it should be entered "https://defined_ip/System-sda.img" "https://defined_ip/vm.cfg";

7. This takes some time, and when it is finished a new VM will be available in the OVMM platform.

This tool takes care of the biggest preoccupations, like file system conversion, or VM configuration files without the need to do anything else then the steps described above.

4.5   Redundancy

So the new adopted infrastructure guarantees the redundancy in terms of servers, preventing that the death of a single server implies the death of the services maintained by it. But there was still a flaw (a big one) in this system. All the storage was saved in one NAS, so if that NAS died, all the services maintained by VMs, would be lost.

To solve this, it was necessary to rearrange both NAS, and then making one NAS the clone of the other. The clone is updated daily, in terms of data it would be better to have a shorter window, however that would cost much in terms of internal band-witch and it would also slow down both NAS (since it were some more operations running all the time), and so a compromise was made, that a day of lost data can be recovered if necessary without too much work.

4.6   New Virtualized Infrastructure

The new infrastructure is represented in Figure 4.4 and we will now explain each item.
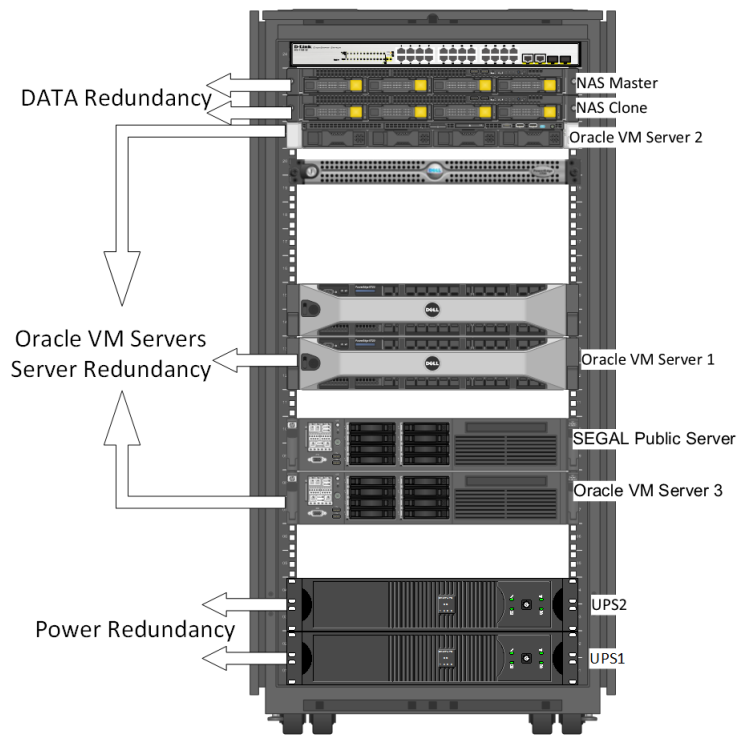
Figure 4.4: SEGAL new arrangement

4.6.1 Hardware description

4.6.1.1 NAS

The NAS have four Hard Drives mounted and were set to a RAID5 configuration. RAID5 is a configuration where one of the HD is configured as a spare. In case of failure in one of the active drives the spare drive takes its place. An illustration on RAID5 can be seen in Figure 4.5.
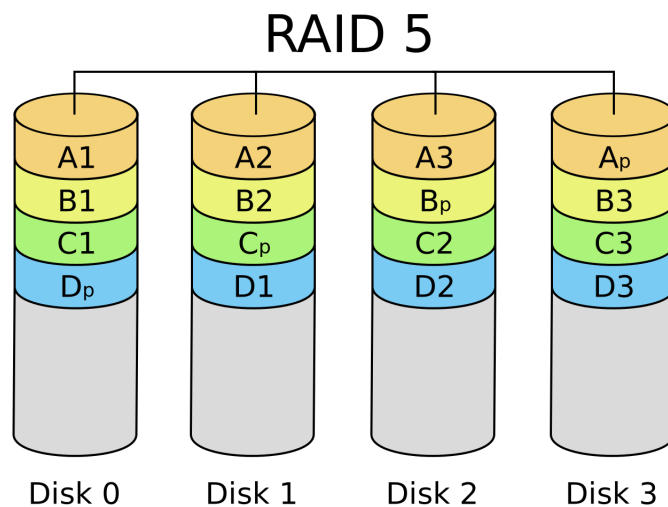
Figure 4.5: Representation of the RAID5

Regarding the NAS hardware:

Figure 4.6: NAS System QNAP TS-420U

Make/Model: QNAP TS-420U (Figure 4.6)

CPU: Marvell - 1.6GHz

Memory: 1Gb

Hard Drives: 4x 4TB Sata

4.6.1.2 Servers

All of the servers involved have a RAID 1 configuration. With the RAID 1 configuration it is ensured that a failure in one of the two Hard Drives will not collapse the whole server.

As can be observed in Figure 4.7 RAID1 is a complete mirror of an Hard Drive to another.
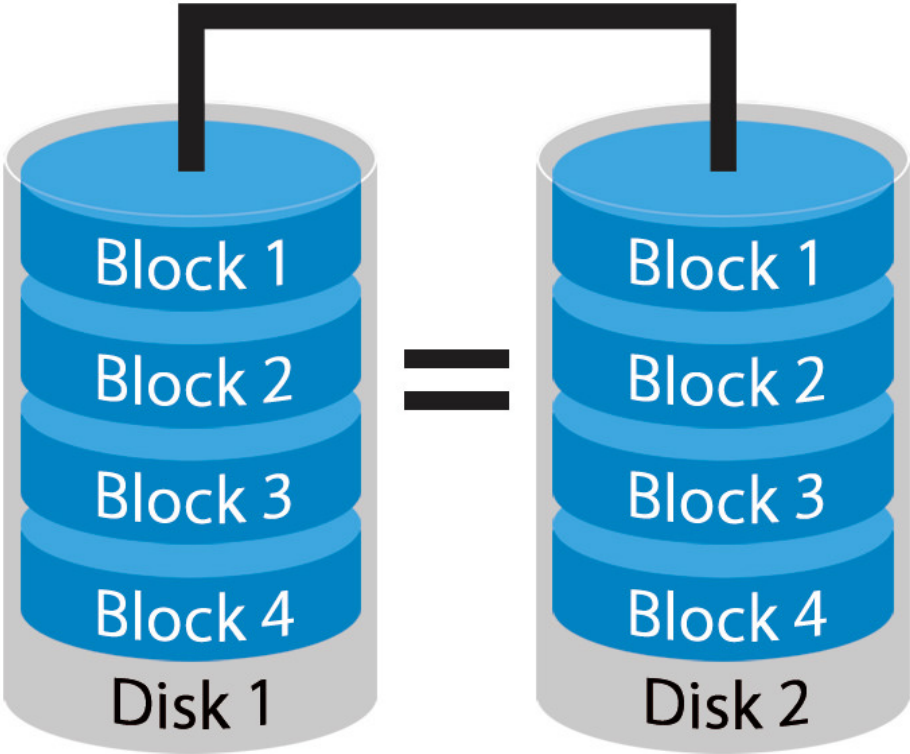
Figure 4.7: Representation of the RAID1

Oracle VM Server 1:

Figure 4.8: Oracle VM Server 1: Dell PowerEdge R820

Make/Model: Dell PowerEdge R820 (Figure 4.8)

CPU: 4x Intel Xeon E5-4607 - 2.6GHz

Memory: 64Gb

Hard Drives: 2x 1TB SAS

Oracle VM Server 2:

Figure 4.9: Oracle VM Server 2: HP ML360 G8

Make/Model: HP ML360 G8 (Figure 4.9)

CPU: Intel Xeon E5-2420 - 1.90GHz

Memory: 8Gb

Hard Drives: 2x 2TB Sata

Oracle VM Server 3:

Figure 4.10: Oracle VM Server 1: HP ML380 G6

Make/Model: HP ML380 G6 (Figure 4.10)

CPU: Intel Xeon E5506 - 2.13GHz

Memory: 8Gb

Hard Drives: 2x500Gb Sata

4.6.1.3 UPS

In terms of hardware involved the only thing missing is the UPS. They are set in parallel configuration, meaning that in case of power loss, the first UPS (lets call it UPS1) starts to provide power to the servers, when that ups is empty, the second UPS (lets call it UPS2) takes its place. It was also configured in one of our server the APC software, that helps to identify when the UPS reaches a certain threshold. Based on that threshold a script was made to shutdown our servers, preventing them from having an abrupt power loss, that can harm them.

UPS1:

Figure 4.11: UPS1: Riello Vision Dual

Make/Model: Riello Vision Dual (Figure 4.11)

Power: 2200Va

UPS2:

Figure 4.12: UPS2: APC SmartUps

Make/Model: APC SmartUps (Figure 4.12)

Power: 2200Va

4.7 Operating System

Another decision that needed to be made was the choice of operating system of the Virtual Machine that would be used during the development and operationalization of the product.
It was set at start that it need it to be a Linux based system. This was mainly because all SEGAL infrastructure is based on Linux systems. But now there were several options. From the most popular Linux Mint and Ubuntu to the more enterprise ones CentOS and RedHat.

The option taken was the CentOS, since it is an open-source clone of the RedHat and that it gives as more support in terms of RAID and that all the packages are very well tested before released for public, and are only released if don't interfere with other functions/packages (e.g.

the official php version in centos 6 is PHP 5.3 that is dated from 2010[Gro01]).

4.7.1 Users

With the Operating System set, it was needed to define the internal structure of the system. First of all were defined the users and their functions:

gpsdata: This user is responsible for collecting all the data for the product. It fetches the GNSS data of the stations from GNSS networks, the Orbits from the JPL servers and the Pressure and Temperature files from SEGAL public server that are uploaded there by IPMA. It also handles a set of secondary functions related with the data collection, like creating log files of the missing that, and the arrival times of the ones successfully collected, that later help perform stats about the system.

realtime: It is the user responsible for generating the processing of the data collected. All the operations that use the GIPSY-OASIS software are performed by this user. It also uses the resultant ZTD of GIPSY and calculate the final product GNSS-PWV.

4.7.2 Folder Structure

Having both the Operating System and the user structure defined, what was left to define was the internal folder structure which largely follows the existing structure for GNSS data that is stored at SEGAL.
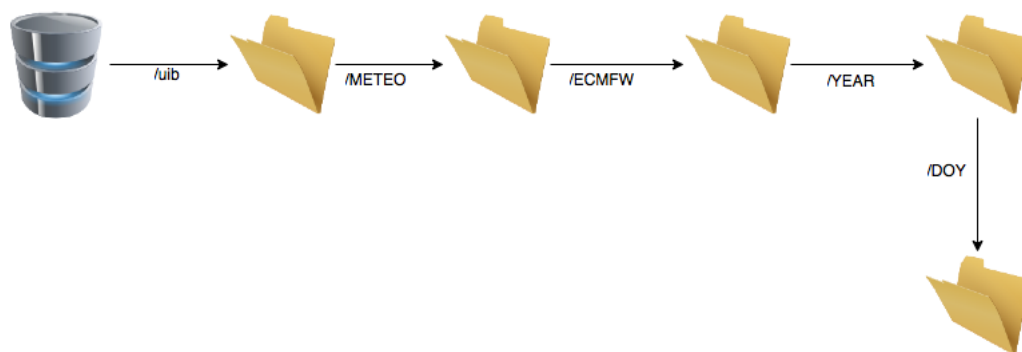
Figure 4.13: Folder structure for the Pressure and Temperature files

In Figure 4.13 it can be observed the folder structure regarding the Pressure and Temperature files where:

"/uib" Stands for the the folder that stores ALL data in the system is located;

"/METEO" Stands for meteorology and it was decided that all meteorological files would be put here;

"/ECMW" Identifies the type of files inside that folder (in this case ECMWF files with pressure and temperature);

"/YEAR" Complete year that prevents data from being replaced, it also makes easy to find old data if necessary;

"/DOY" It translates to Day-Of-Year and together with the above folder it makes very easy to organise the data.
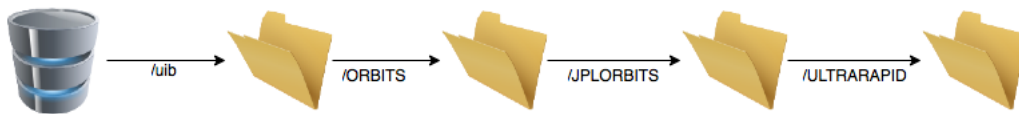
Figure 4.14: Folder structure for the Orbits files

It is represented in Figure 4.14 the structure that holds the Orbits files and where:

"/uib" Stands for the the folder that stores ALL data in the system is located;

"/ORBITS" Identifies that inside are Orbits;

"/JPLORBITS" Identifies that inside are Orbits provident from the JPL;

"/ULTRARAPID" Identifies the kind of Orbit.

In this case it is not necessary a Year and Day-Of-Year Folders because the ultra-rapid orbits are deleted after being used. If for some reason it is necessary to calculate the GNSS-PWV after a few days, other, more precise Orbits, will be available.

Finally it can be observed in Figure 4.15 the structure for all the NUVEM operational system:

"/dropbox" Contains the last 30 hours GNSS files for each Station;

"/history" After the 30 hours window the GNSS files are compressed and putted here organised by Year, Day-Of-Year and Site;

"logs" It contains all the logs of the system, delay in terms of files, files missing, etc.;

"/preprocess" Temporary folder used to store generated GNSS files combining 30 hours (used in the GIPSY processing);

"/stats" Holds all the statistics of the system;

"/PLOT" It contains the last generated plot of the GNSS-PWV values;

"/SOLUTIONS" Inside this folder are located the calculated solutions;

"ZTD" It keeps the calculated ZTD files organised by Year, Day-Of-Year and Site;

"PWV" It keeps the calculated PWV files organised by Year, Day-Of-Year and Site;
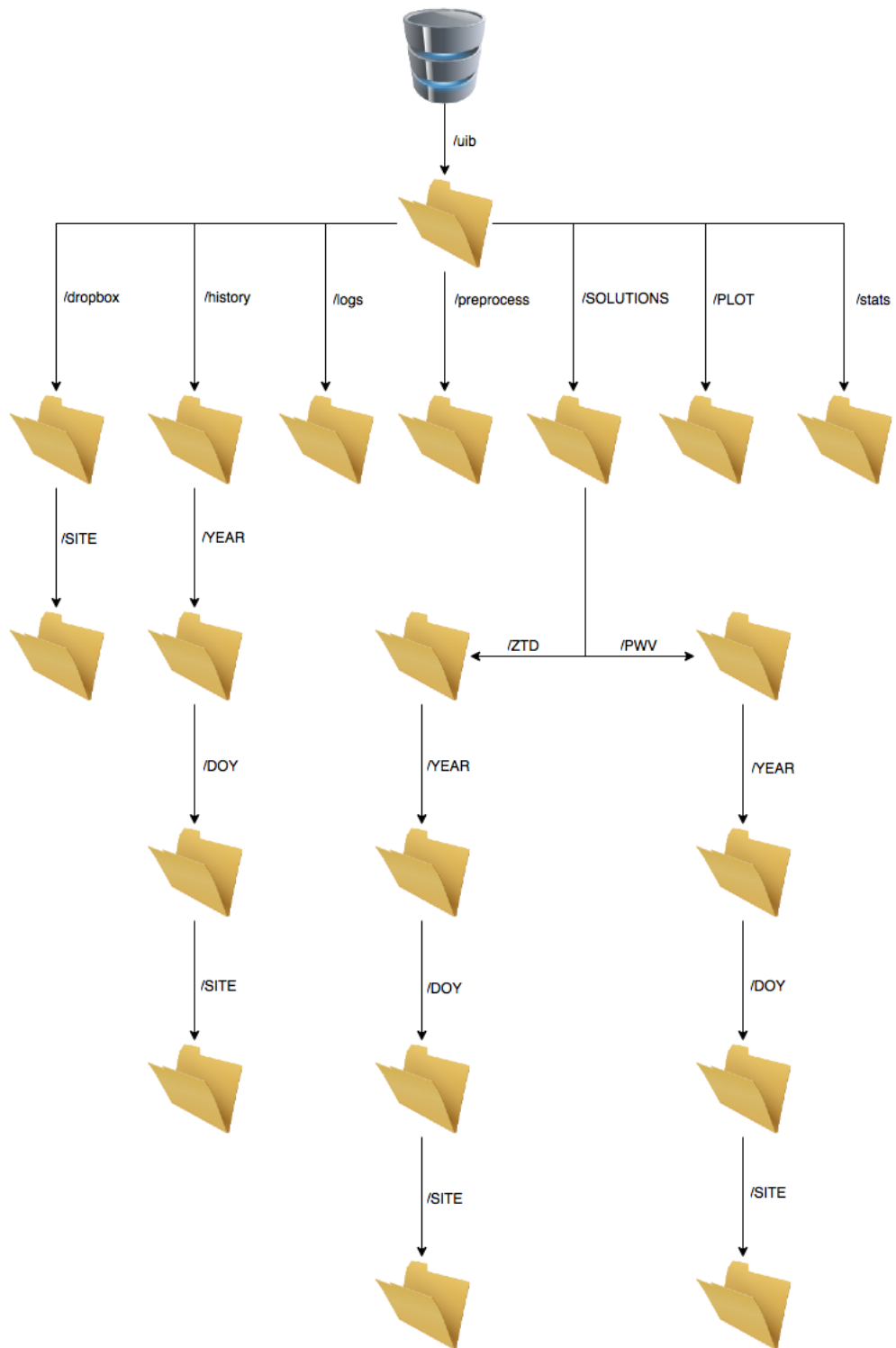
Figure 4.15: Folder structure for the NUVEM system

4.8 Summary

In this chapter an overview of the changes made to SEGAL infrastructure, in order to provide a more secure system for the entire SEGAL environment, including the computational resources dedicated to the NUVEM project, was made. It was also made a description of the options taken in account for this changes, and finally a descriptions on the internal structure of the NUVEM server was done.

Chapter 5

Near Real Time Data Acquisition

5.1 Networks

At the start of the project it was necessary to decide which GNSS stations would be important to use. Of course all stations in Portugal would be of interest. In Portugal there are two GNSS networks, ReNEP and SERVIR, both are owned by governmental agencies.

ReNEP: Is maintained by the Territory General Management (DGT - Direção Geral do Território). ReNEP has a set of 47 GNSS stations with a good spatial homogeneous distribution. This network as a public FTP server where the GNSS data can be collected.

SERVIR: Is the Portuguese Army Network. It is composed by 30 stations. Is a closed network. University of Beira Interior (UBI) through SEGAL have been holding a station, and in exchange they provide their data to SEGAL. After a meeting between the UBI, IPMA and the Portuguese Army, it was possible to establish a cooperation agreement so that the GNSS data could be used in the ambit of the NUVEM project.

Both networks combined they provide a set of 77 GNSS stations distributed all over the Portuguese territory.

But not just the Portuguese stations are of importance, since storms, and clouds move, it was also important include in the processing the Spanish stations close to Portugal. With this in mind a fictional barrier was made in the vertical axis of Madrid (in a 2 Dimensional Map like shown in figure 5.1). With this it was easy to identify the stations of interest. In Spain instead of a more centralized approach for the networks, each region has its own center.

ANDALUCIA: This network has a set of 17 GNSS stations. The data is open to the public, through a set of formularies via the website (www.ideandalucia.es).

CASTILLA: The GNSS network for this region is composed by 28 stations. The data is available by an open FTP server.

EXTREMADURA: In the Extremadura region the GNSS network has 11 stations. The data is available from an open FTP server.

IGN: Is the only network that has stations all over Spain. It is owned by the Spanish National Geographic Institute (IGN), to this case their network has a set of 20 stations of interest.

Having all of those institutions in account there are currently 76 GNSS stations in Spain that will provide data to the NUVEM project.

Figure 5.1 shows the map of the GNSS stations, used, in both Portugal and Spain.
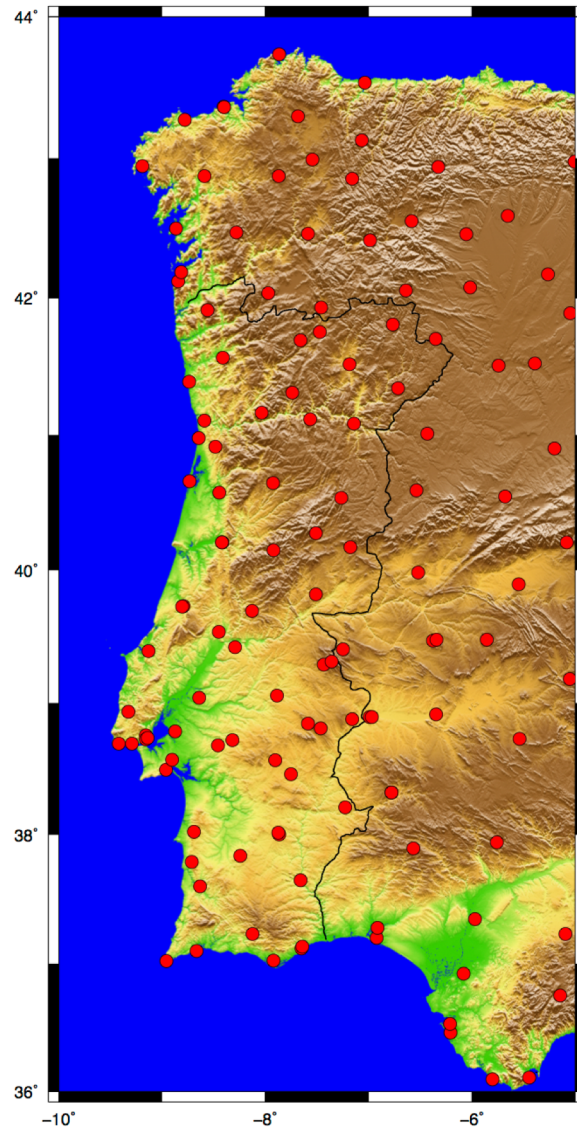
Figure 5.1: Map with the Sites used in the NUVEM project.

5.2 Technologies used

In this section a look at the technologies and software used during the development of the near real time system will be taken.

5.2.1 TCSH

The first step was to take a look at the requirements and based on that, choose an appropriate programming language. The NUVEM research team decided to implement the system using a Linux scripting language, instead of a programming language. The argument is that a Linux shell scripting language, is lighter than a programming language, and since there isn't the need for any advanced tools, like advanced data structures, or even the need to use a database, the choice was made.

It also needed to decide which Linux shell scripting language, the choice, was TCSH. It is based in the C shell(csh), and the main reason here is the fact that this scripting language has been used at SEGAL for many years to develop processing routines. So not just that integration is easier, in the future if there is a need for changes, it will facilitate the SEGAL team.

5.2.2 Java

During the development, a problem was found (discussed below). One of the networks, just allow data download through a set of html forms. To automate this task, Java[oqu15] and two external libraries were used. Java is currently on of the most used programming languages in the world and it is owned by oracle. The reason here was the vast experience in Java by the developer. The external libraries used were:

HtmlUnit: Is a GUI-Less browser for Java[BGA16];

JSoup: Is an HTML parser[Hed09].

5.2.3 Linux Tools

Since a Linux shell scripting language was in use, it was necessary, to use a set of Linux tools. Some were more simple and common like the commands "cat, grep, cut", but there are a few others more complex and more important for the matter at end.

crontab This is the Linux task scheduler.

awk Is a programming language by it self. It completes the shell scripting very well because it is very easy to parse files, and strings, and make calculations with it.

RSYNC is a software that allows a complete synchronization between folders (and their structures). This synchronization can be inside the same machine, or like in our case between different machines.

5.2.4 TEQC

Another tool used is TEQC[UNA16a] (Translation, Editing and Quality Checking). It is freeware, that has multiple appliances regarding GNSS data. It has been developed and is maintained by UNAVCO[UNA16b].

5.3 System Architecture

In terms of the system development, it was divided in three parts, Data Collection, Data Processing, Results/Stats Generation and upload to IPMA. Figure 5.2 describes the developed system.
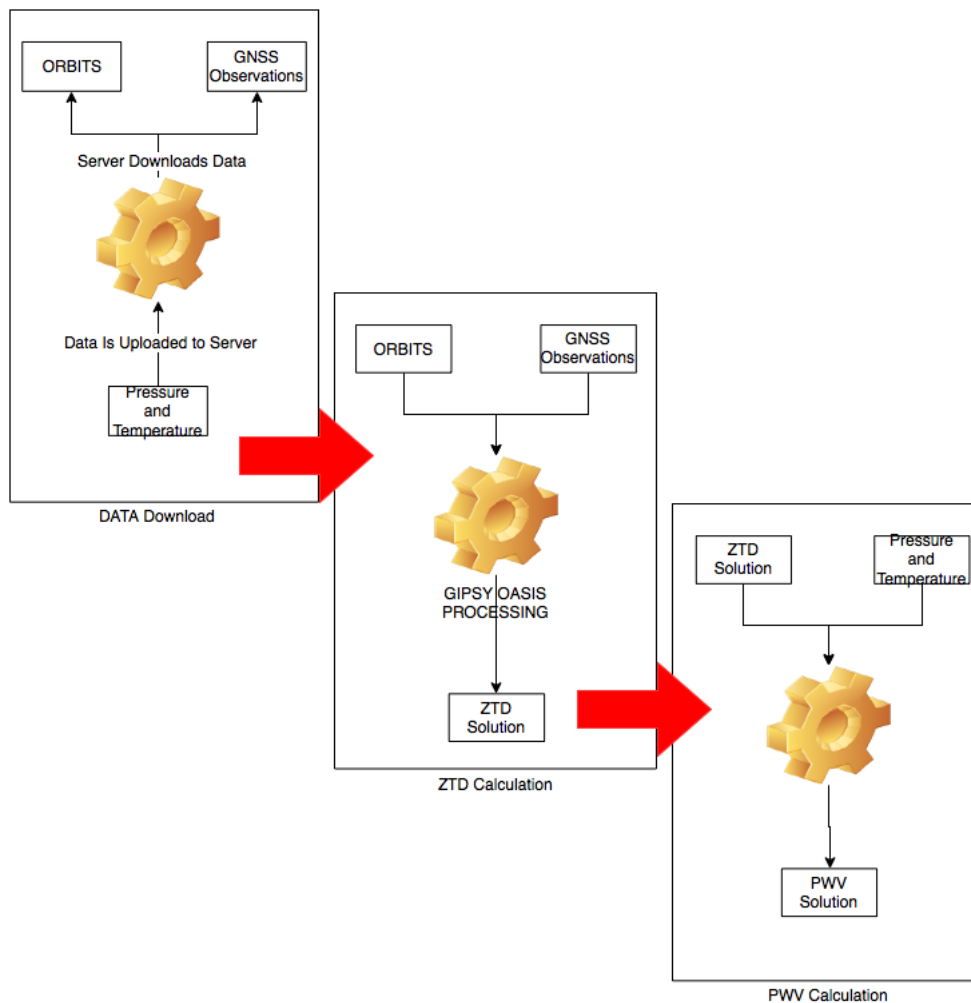
Figure 5.2: General Diagram of the System.

Bellow is a detailed look at all the parts of the system implemented.

5.3.1 Ultra Rapid Orbits and Clock Acquisition

It was necessary to acquire the the Ultra Rapid Orbits and Clocks. As discussed earlier in Chapter 2, these Orbits and Clocks are downloaded from JPL.

JPL provides this data through an open FTP server (ftp://sideshow.jpl.nasa.gov/pub/JPL_GPS_Products/Ultra/) and organizes the data in yearly folders (i.e. the folder "2016" contains all the orbits and clock files of the year 2016). Inside that folder, the file names respect a certain structure "YEAR-MM-DD..." where "YEAR" is the complete year, "MM" stands for month in two digits, and DD stands for day, also represented with two digits. for each day there are several files, and as said before, the ones we are interested are updated during the day (every hour). To help find out the update, JPL creates in its FTP an empty file that indicates the last updates (i.e. let us assume it is 16:00 UTC, the empty file with the name "2016-05-16.14.00.done" indicates that the orbits and clock files have already the update hours until 14:00 UTC. Remember that there is a two hour delay.) Figure 5.3 shows a diagram for this part of the system.
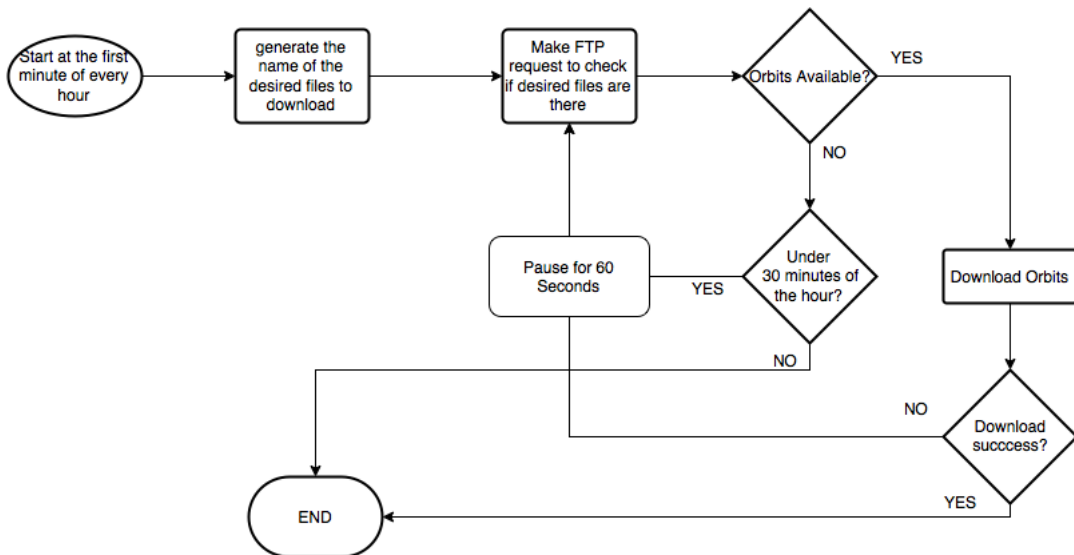
Figure 5.3: Diagram of the Orbits and Clock files download.

This script starts at the first minute of every hour (using crontab for that), the script generate the name of the desired files, after that the script makes an FTP request in search of the "done" file for the desired hour. In here the script if the orbits are available (if the "done" file was available) the script attempt to download the orbits and clock files, if it has success the script terminates successfully. On an alternative path if the orbits are not available the script enters in a loop where a 60 seconds pause takes place, and then retry the download. If the script is in the loop, at any iteration it also verifies if the minute is bigger than 30. That is the window defined for the script to end. If by that time the orbit is not there, it is not important anymore.

5.3.2 Pressure and Temperature

Regarding Pressure and Temperature files, these are provided by the project partner IPMA. Because of their internal structure they wont allow SEGAL to collect the data. In this case the adopted approach was, that IPMA had FTP access to SEGAL's public server and uploaded the data to there, and then the NUVEM Virtual Machine would collect it there. This case can be seen in Figure 5.4.
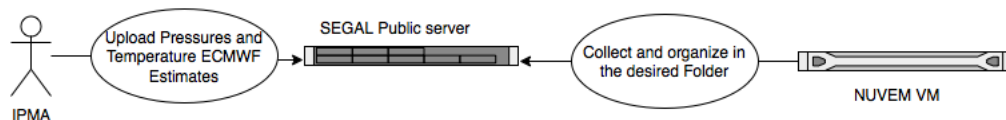
Figure 5.4: Diagram IPMA uploading files to SEGAL.

These files are uploaded twice a day to SEGAL:

At 7:00UTC ECMWF estimative (one file for each station) are uploaded, these contain data from the same day from 12:00 to 23:59.

At 19:00UTC ECMWF new files are uploaded, containing data from the next day from 00:00 to 12:00.

39

In SEGALs public server the data is stored as it can be seen in Figure 4.14.

After that it is necessary to collect the data from SEGAL's public server to the VM. A detailed diagram can be seen in Figure 5.5.
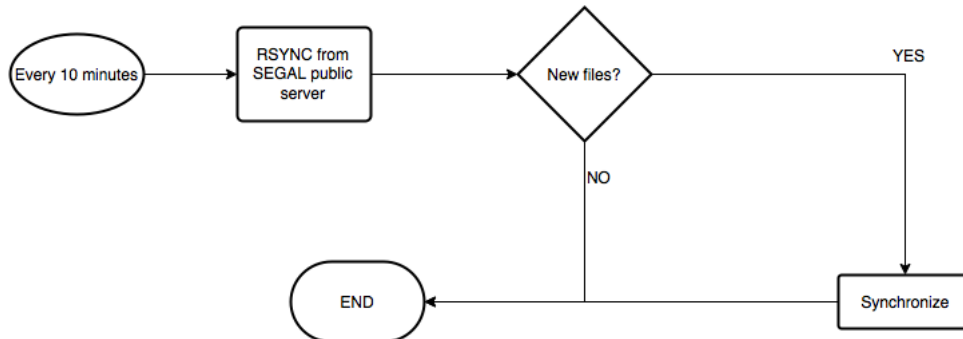
Figure 5.5: SEGAL internal synchronization of the Pressure and Temperature files from the Public Server to the NUVEM Virtual Machine.

This script runs every 10 minutes, it calls the RSYNC software, if it has new files, it synchronizes, and terminates, if it doesn't have anything to synchronize it terminates immediately.

Once the synchronization is done there is still another step missing. It is necessary to convert these files in a format that is compatible with GIPSY. To do that the UNAVCO's TEQC software is used. For the conversion of the files, a script was made, and its flowchart can be seen in Figure 5.6.
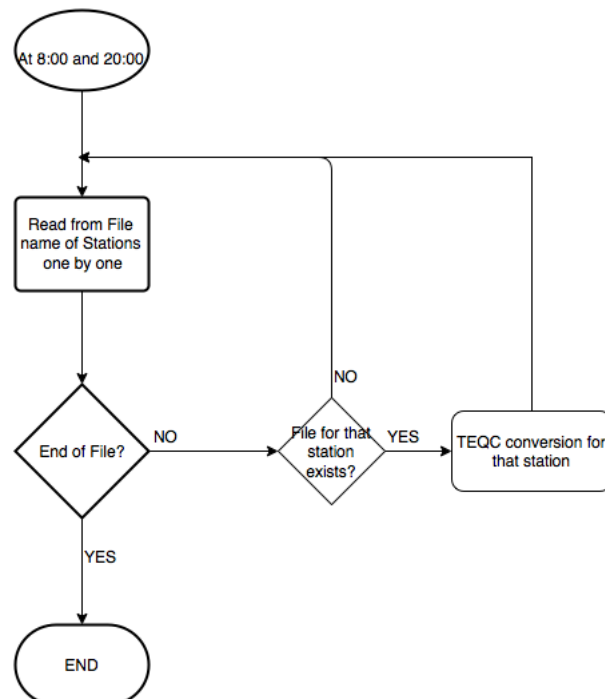
Figure 5.6: Pressure and Temperature file conversion.

The script runs twice a day, at 8:00 and 20:00 (one hour after IPMA has uploaded the original

files), after that the script reads the station names one by one, checks if the Pressure and Temperature file is already in the system, if it is runs the TEQC conversion, and moves for the next station, until it reaches the end of the station list. If it is not, moves for the next station in the list. After this last step the files are ready to be used in the ZTD to PWV conversion.

5.3.3 RINEX files acquisition

Since gipsy work better with a 30 hour window file instead of just the desired hour, the script developed needs not just to download the files for the last hour, but also, if they are missing, files up to 30 hours (instead of just the 30 hours, it was decided to download up until 48 hours). Every network has a set of files that help the system workflow:

stations_network: It contains the station list for the network;

stations_network_off: It contains the stations of the network that are offline (more on this below);

downloads_missingNETWORK.log: This log contains the downloads missing in the last 48 hours for that network;

download_bad_NETWORK.log: This log contains files that weren't download even after the 48 hour windows;

For each network a separate script was developed (they are almost the same with some minor changes, because of each provider rules, except for the ANDALUCIA network). In Figure 5.7 the script flowchart is represented.

5.3.3.1 RENEP / IGE / Castilla

Figure 5.7: Flowchart GNSS data download.

The script starts at the first minute of every hour. The first step is to clean the log (down-loads_missingNETWORK.log) for records older than 48 hours, then it reads the stations, one by one from the "stations_network" file, then it generates the name of the desired file for download from that station, and after it has the name, tries to download, if it has success, it creates

42

the flag (that indicates the processing Daemons, more below, that the station is ready to process) file and moves for the next station. If it fails to download, write the missing file to the "downloads_missingNETWORK.log" log. When it has tried all the stations it is time to try and recover the files listed in the log. For a faster approach the script makes up to a maximum of 40 downloads requests at a time(this threshold was defined after a few problems with the number of requests per minute done to some providers). If it reaches the limit, it waits for the current downloads to end, and then proceed for the files still in the log. As the downloads succeed the log is cleaned. The script terminates if the log file is empty, or if the minutes (that are checked at any pass) are higher than 50 (this is to let the script finalizes all the sub-process) in order that only one iteration of the script is running at the same time (because it can damage the files structures if more than one processes are using them).

5.3.3.2 SERVIR / Extremadura

These two networks use the same system that is described for the previous networks, and that is shown in figure 5.7. The main difference is that for these networks, for different reasons, SEGAL wont collect the data, it is uploaded to SEGAL's public server. Very similar to the scenario described in Figure 5.4.

SERVIR: Regarding SERVIR, being a military network, for security reasons they wont allow any one to connect to their servers. As stated before a cooperation agreement was established for this network.

Extremadura: This Spanish network is a different case. After a few weeks of the system being operational we were contacted by the System Administrator, stating that even our threshold of 40 requests a minute was too much for their system. As an alternative he suggested that he uploaded the files to our server as soon as they are available, that option was accepted at SEGAL.

5.3.3.3 Andalucia

Andalucia is a very different case from the others. They don't offer an FTP connection. The data needs to be downloaded through their website, through a set of formularies. To overcome this obstacle it was resorted to Java, and a few of its libraries, HtmlUnit and JSoup. Not just that it is not possible to download an hourly file. The only option available is to download all the hourly for the desired day. So if the script is looking for the 16:00 file, it has to download a ZIP archive with all the hourly files of the day, uncompress and see if the desired file is there or not. In Figure 5.8 is represented this script flowchart.
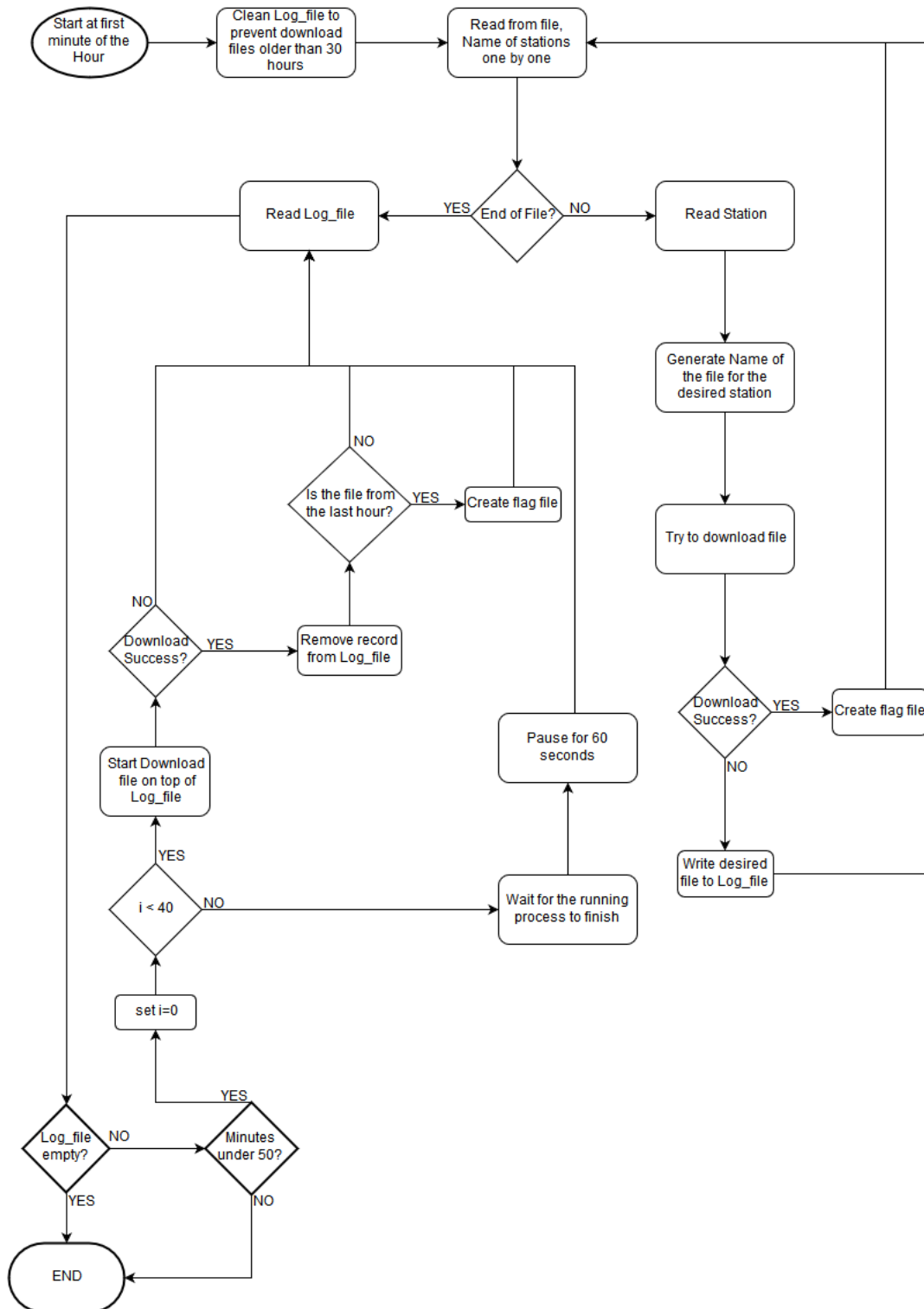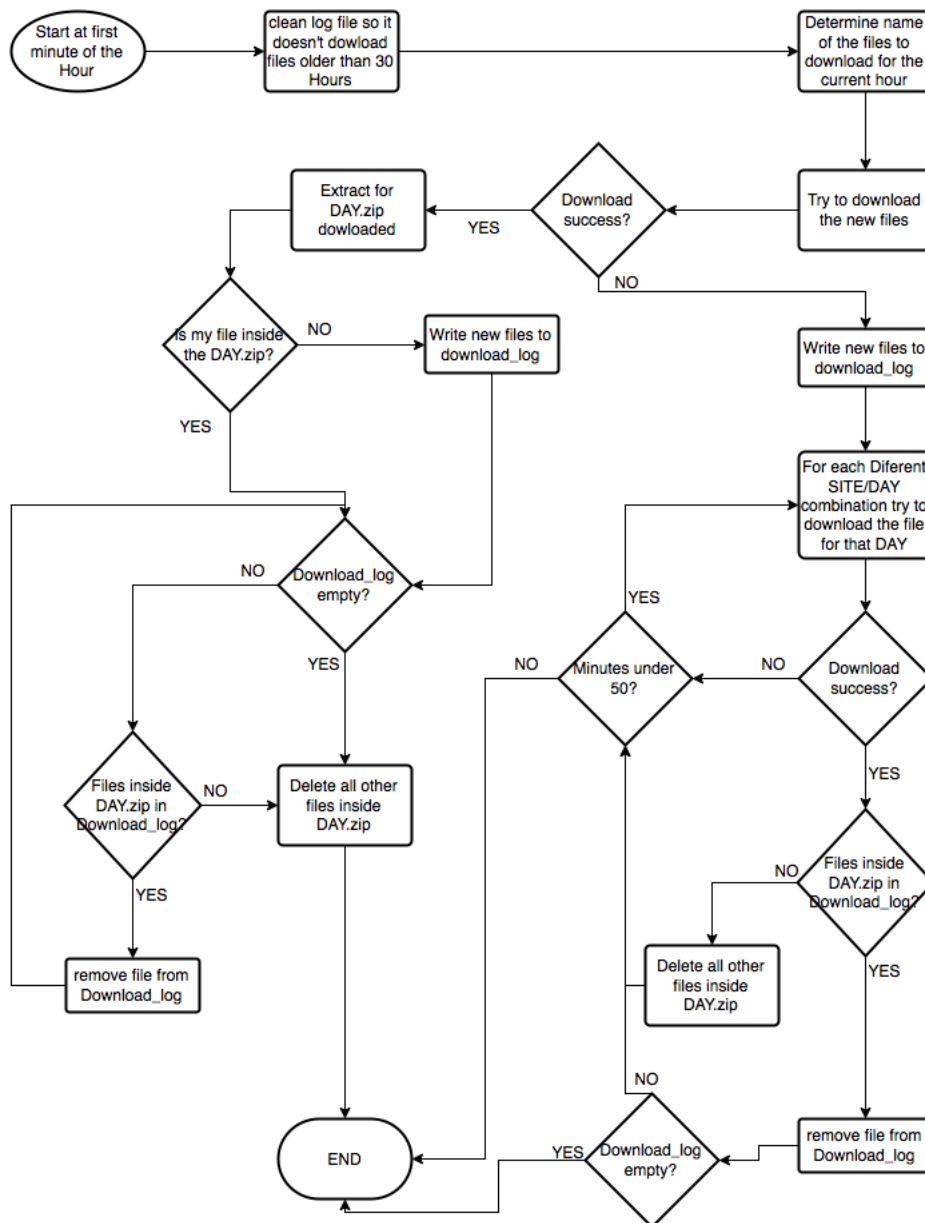
Figure 5.8: Flowchart GNSS data download - ANDALUCIA.

The script starts at the first minute of every hour. The first step is to clean the log (downloads_missingNETWORK.log) for records older than 48 hours, then it reads the stations, one by one from the "stations_network" file, then it generates the name of the desired file for download from that station, and after it has the name, tries to download, if it has success, it extract the ZIP archive, and verifies if the desired file is inside, if it is it creates the flag, if not write that file to the log file. After it verifies if the log is empty, if not, sees if any other file of the ZIP was in the log, then it deletes the remaining files contained in the ZIP. After trying all the most recent files, it checks if the log file is empty, if not it downloads the ZIP archives for the desired days, and search for the missing files. It keeps this loop until the log file is empty or the minutes are higher than 50.

5.3.4 Processing

After collecting all the necessary items to generate the GNSS-PWV it is necessary to create it. For that there are two steps. For each step it was developed a Daemon that monitors something. The first step is to check the flags that mark the stations that have updated GNSS data and compile a 30 hour GNSS files (this data span corresponds to the orbits data span). In figure 5.9 is represented the script that does this part.
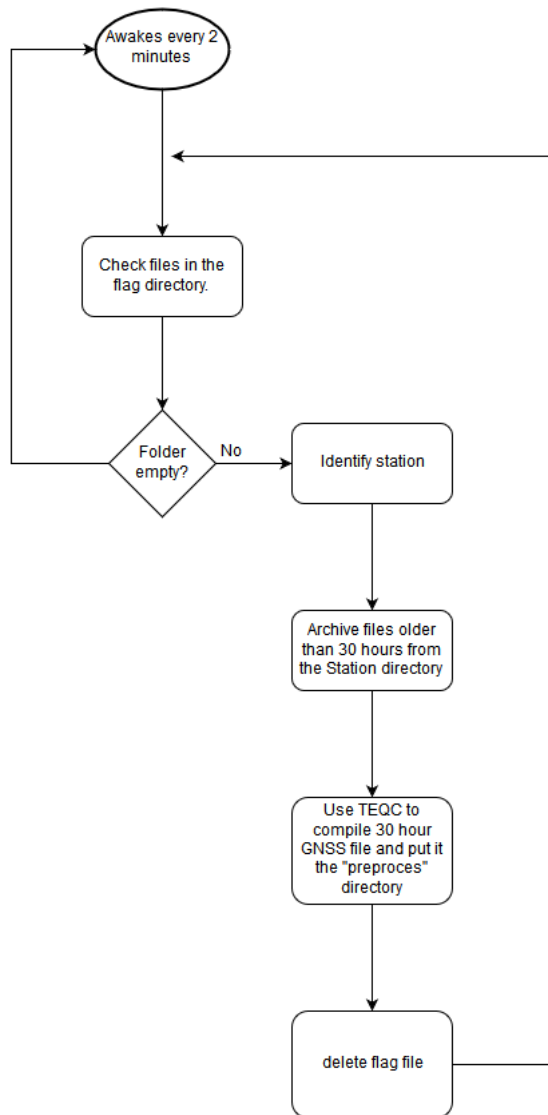
Figure 5.9: First step of the processing scheme.

With crontab, we start the Daemon at every 2 minutes, if for some reason the Daemon has terminated, it will start, if everything is running has expected, this new Daemon will die almost immediately. After that the script looks for the folder with the flag files (/uib/nuvem/dropbox/flag) and read the names of the stations ready for processing from the file names inside that folder one by one. After reading the station name, the script goes to the station folder (/uib/nuvem/dropbox/SITE/) and archives files older than 30 hours for that station in the appropriated folder (/uib/nuvem/history/YEAR/DOY/SITE/). The next step is to use TEQC to compile the 30 files from the last 30 hours into one file, and put it in the preprocess folder (/uib/nuvem/preprocess/),

45

after that it deletes the flag file, and moves to the next station. If the folder is empty (there is no files left to compile) it pauses for two minutes before checking the flag folder again.

The second part is to process the generated 30 hours GNSS data files, using the orbits and clock files, with the GIPSY software. That gives us the ZTD. After that we use the ZTDs obtained, make use of the Pressure and Temperature estimates and calculate the PWV values. In figure 5.10 is represented the script that does this part.
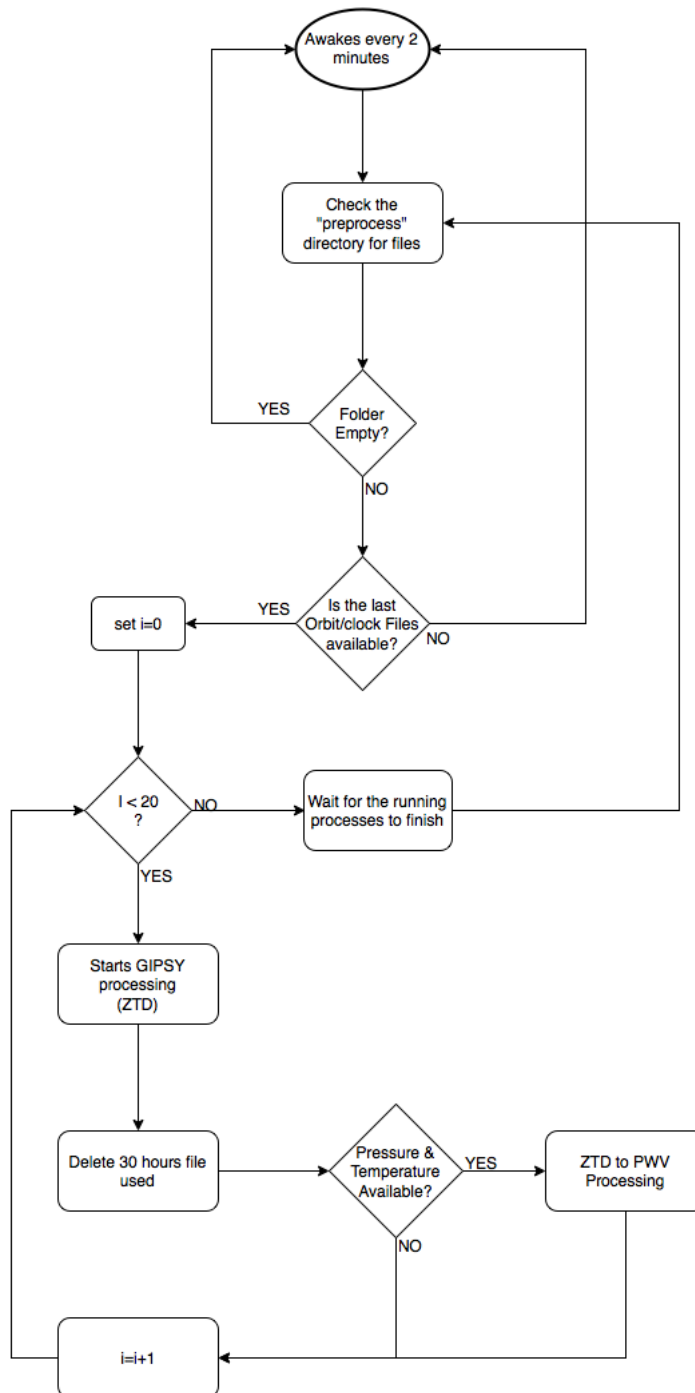
Figure 5.10: Second step of the processing scheme.

In terms of script execution, the idea is the same used in the previous script. When it starts

this script checks the files inside the preprocess folder (/uib/nuvem/preprocess/), if the folder is empty, it pauses for 2 minutes before restarting, if it is not empty, it checks is the necessary Orbit and Clock files are available. If they are not available, the script pauses for 2 minutes and restarts. If the Orbit and Clock files are available and there is files to process, the script starts up to a maximum of 20 GIPSY processes in parallel. If the maximum is reached the script waits for all the process running to finish and then rechecks the folder for remaining files.

5.3.5 Graphic Generator

At the end of the processing, there is a script that runs at SEGAL to generate a plot with data from the last hour. That plot is aimed to be seen by the public, in the NUVEM website (http://nuvem.di.ubi.pt/) that was also developed during the course of this dissertation. In Figure 5.11 it is an image of the website. The plot is a movie from the last hour, with increments of 5 minutes.
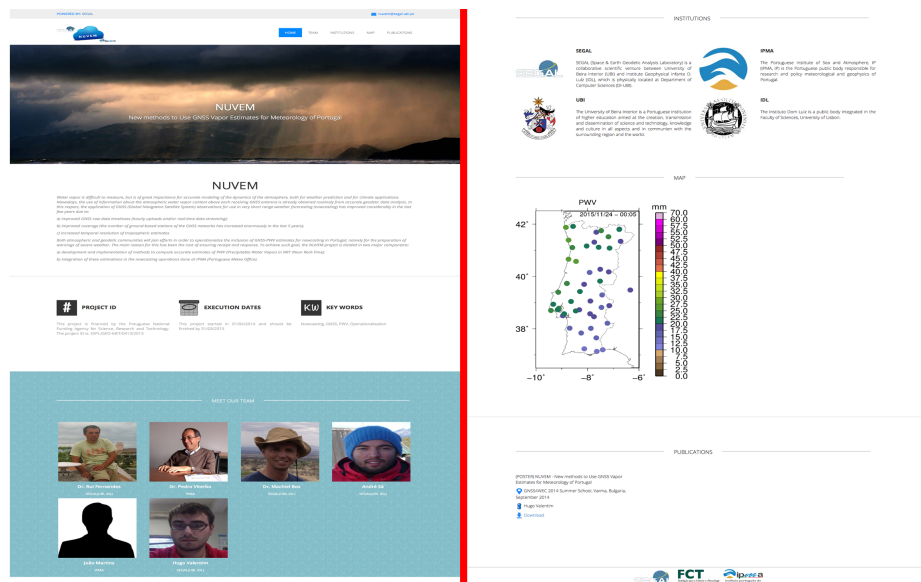
Figure 5.11: Image of the NUVEM project website.

To generate the images that compose the movie, it is used a script, developed by Dr. Machiel Bos, already used at SEGAL for similar purposes for several years, just some minor minor changes were necessary in that script. The script receives as argument the following arguments:

year: The year of the data to be processed (since the data has a delay of two hours, it is possible that the year has changed);

year_solution: The year of the solution, the solution is named after the current date, not on the data it contains, so the first solution of the first day of 2016, only contains data of 2015;

day: Like in the year case, this is the day of the data to be processed;

day_solution: The day of the solution;

second: Second of the day to plot;

hour_solution: Hour of the solution.

The arguments received, are mainly to deal with the folder structure, and with the names of the solutions (that differ from the data they contain). As an output of this script it is obtained a static image with the map, with the stations, and respective PWV representations for the requested second. In order to be able to produce the movie it is necessary to call that script 12 times each hour (incrementing the seconds value by 300 (5 minutes)) and in the end combining all of them into a gif movie. The process to generate the movie can be found in Figure 5.12.
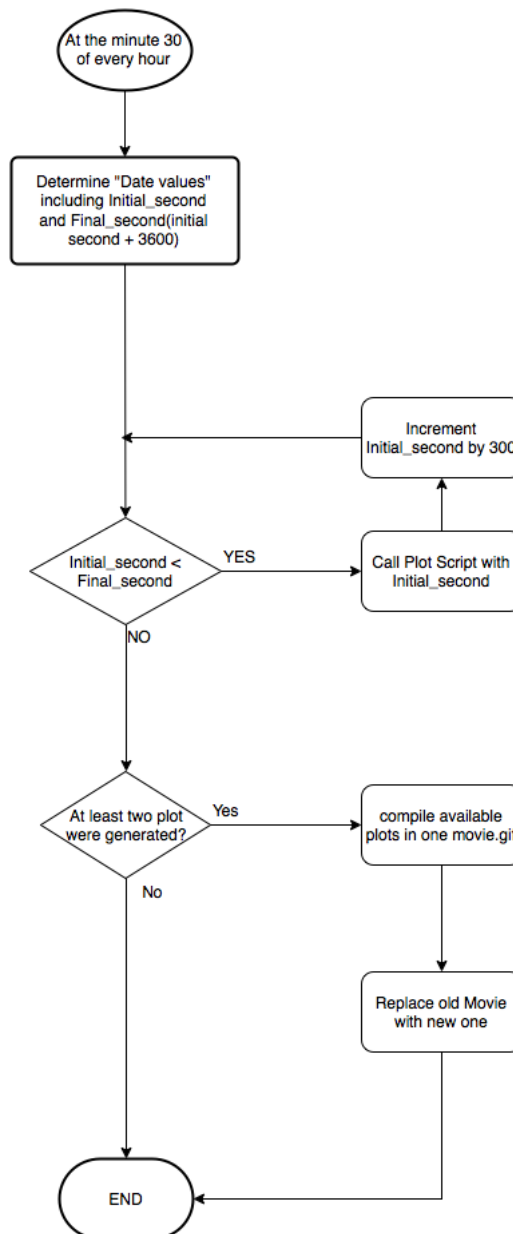
Figure 5.12: Movie of an hour generation diagram.

Since the movie is generated for public demonstration only, it was decided by the team to generate at the minute 30 of every hour with the available data (it can occurs the movie only has info about 1 station, or that it has only data from a 10 minutes window instead of 1 hour). The script starts at minute 30 of the hour, then it determines a set of variables based in date values, including the initial second of the desired hour (i.e, initial second for hour 2:00 is second 7200)

and the final second for the same hour (initial second + 3600 seconds). After it calls Dr. Machiel plot software until the initial second and final second have the same value (at each pass the initial second increments 300 seconds). When that condition is reached the script verifies if at least two plots were generated (meaning that there is at least data from a 10 minutes window), if not, the script terminates, if there are more than 2 plots, the script compiles the available plots in one movie.gif, and finally replaces the old movie with the updated one.

5.3.6   Auxiliary Scripts

To optimize the system it was necessary to develop some auxiliary functions (scripts).

5.3.6.1   Clean Stations

It is possible that for some reason some stations became offline for several days or weeks. To prevent the system from keep trying, in vain, to access data from that stations a script was developed (figure 5.13).
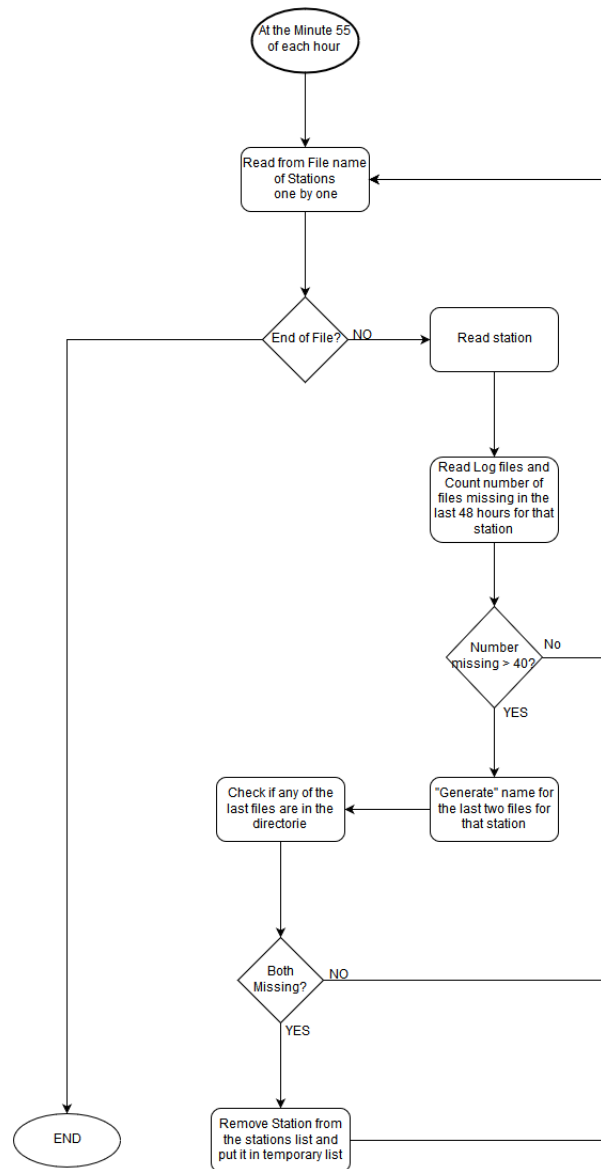
Figure 5.13: Clean Stations auxiliary script diagram.

The script runs at the minute 55 of each hour, it reads the stations, one at a time, from the "stations_network" file (one network at a time), then read the log file for that network, and count the number of files missing for that station, if there are more than 40 files missing (the maximum that can miss are the 48 representing the window given for download) it generates the names of the 2 more recent files of that stations and checks if any of those 2 is present in the system. If both are missing remove the station from the "stations_network" file and put it in the file "stations_network_off". If at least one is in the system it moves for the next station.

5.3.6.2 Retest Stations

Since it is possible to put stations on hold. It is also necessary to check if those stations are back online (figure 5.14).
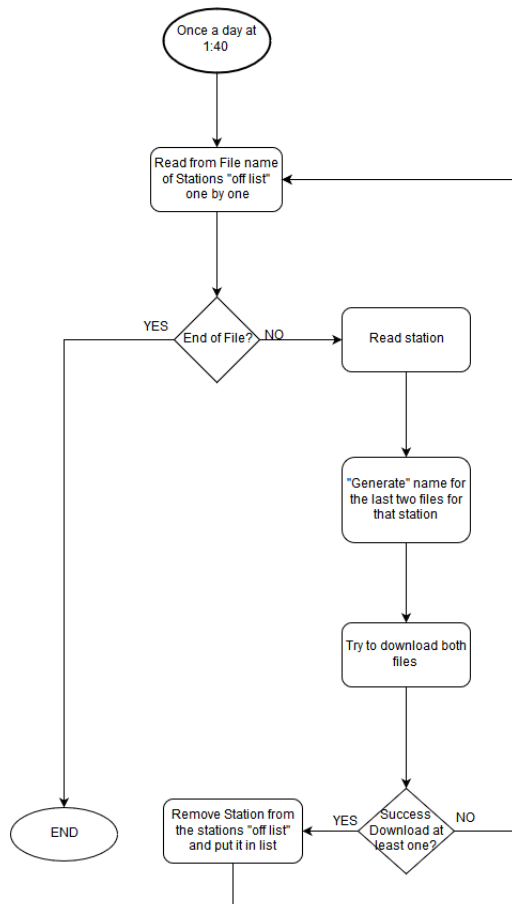
Figure 5.14: Retest Stations auxiliary script diagram.

The script runs once a day at 1:40, it reads the stations, one at a time, from the "stations_network _off" file (one network at a time), then it generates the names of the 2 more recent files of that stations and attempts to download. If it has success in at least one of them it removes the station from the "stations_network_off" file and put it in the file "stations_network".

5.3.6.3 Clean 30 Hours

As seen before, when the GNSS data is downloaded the files older than 30 hours are compressed and archived. But if for some reason the station is off for a lot of time that data will not be on the place they are supposed to be (figure 5.15).
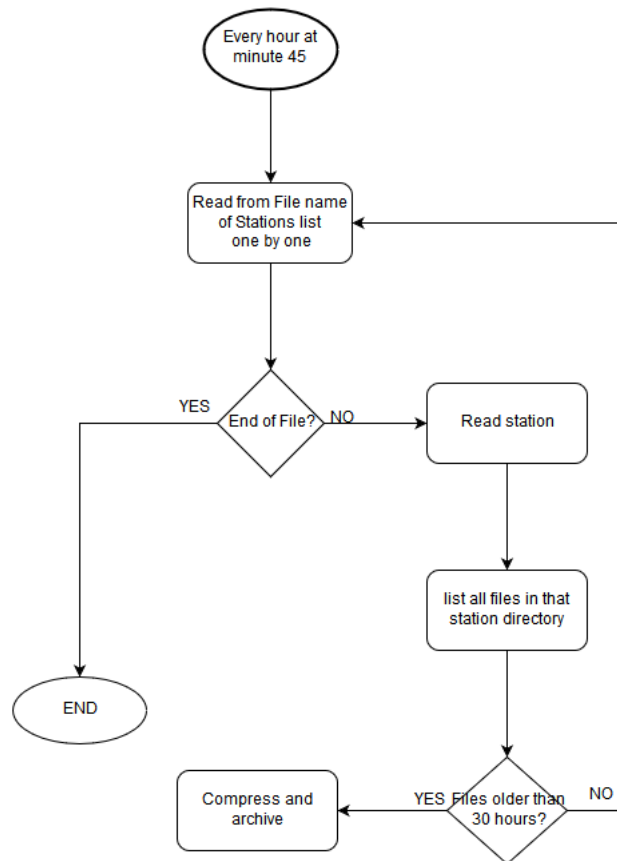
Figure 5.15: Clean 30 hours auxiliary script diagram.

At minute 45 of every hour, read the stations name, from a file "stations_all" that contains a list of the stations of all the networks, one by one. Check the station directory (/uib/nuvem/dropbox). If there are files older than 31 hours, compress them and archive. If not, moves for the next station.

5.4 Summary

In this chapter it is taken a look on the GNSS networks used in the NUVEM project. It is also described the system architecture, and the tools used, and some reasoning on why they are used.

Chapter 6

Results & Data Validation

Several tests were performed in order to evaluate the reliability of the implemented system. These tests focused on two main areas:

1. the reliability of the internet network in terms of transferring the necessary data (observations and orbits) timely.

2. the quality of the estimated solutions, in particular to evaluate the influence of data missing.

6.1   Network Reliability

The first aspect to assess was the the network reliability. The questions aimed to be answered in this sections, are:

How many files were missing for each station?

What is the average delay for each station?

Thus, it the effect of availability of the observations, both in quantity and time, were investigated. In order to answer these two questions data from 1/10/2015 until 20/05/2016 (a total of 231 days) was used to compute some statistics .

Regarding the Castilla network, unfortunately there are no data. The reason for this is that during an initial stage of the development, our download script caused a crash on their FTP server. After a few discussions with them to solve the problem, they demanded a maximum of 40 requests per minute. Although this demand was met, after a few weeks, they told us that it was still to much, and they blocked SEGAL's public IP. This situation was solved a few days only before the delivery of this dissertation.

6.1.1   Data availability

The total number of expected files for the studied period were 5544 (24 hourly files * 231).
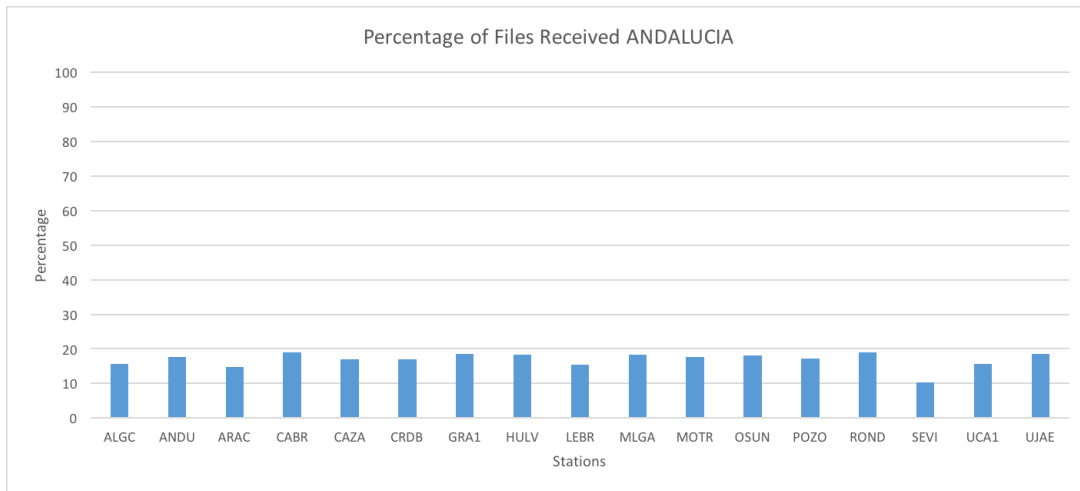
Figure 6.1: downloaded files per station (ANDALUCIA network)

Figure 6.1 clearly shows that the ANDALUCIA network was not very reliable. During the tests conducted it was evident that the repository is not updated every hour (like most networks are). Sometimes the repository is just updated once a day. It is believed that this is the cause of such bad results in this test. The average percentage of received files for this network is just of 16,9%.
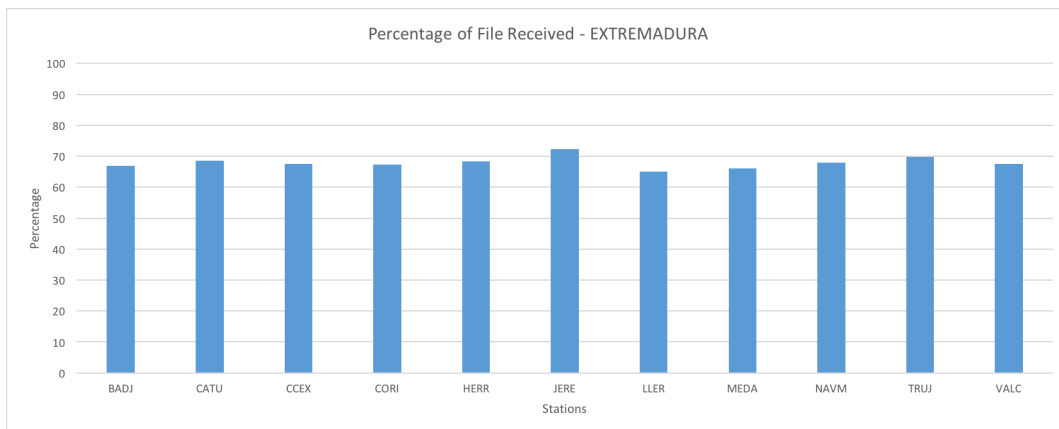
Figure 6.2: downloaded files per station (EXTREMADURA network)

For the EXTREMADURA network the average percentage of received files is 68% (Figure 6.2). Even though this percentage is much higher than the previous network, it is not ideal yet. There is however a reason that can justify the results. During this period there was a problem in the server of the provider, and after they correct it, the upload rule to SEGAL was not restored at first. This was only corrected after 44 days.
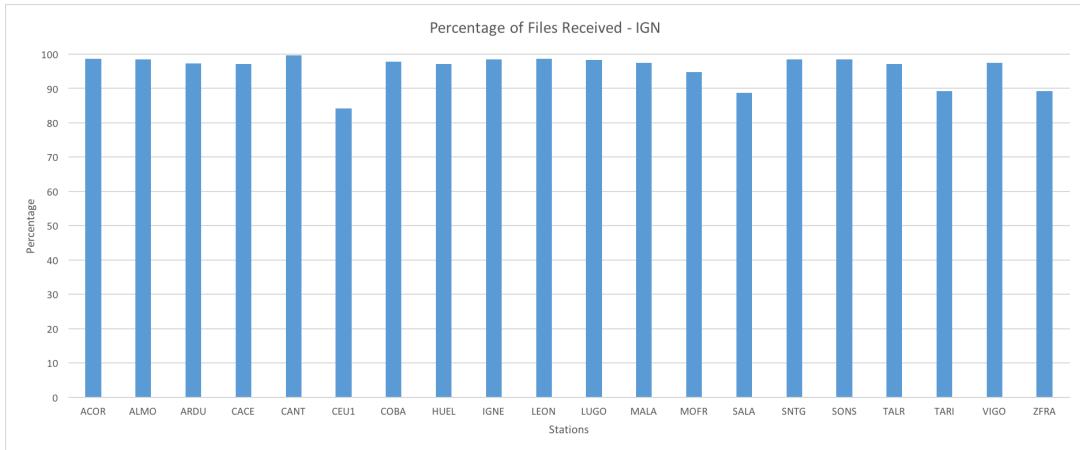
Figure 6.3: downloaded files per station (IGN network)

The IGN network is very reliable as shown in Figure 6.3. The average percentage of received files is of 95,8%.

The ReNEP network is also very reliable, with an average percentage of received files of 88,8%. This percentage is damaged by the fact that 2 stations where off-line during the whole time of the test, as shown in Figure 6.4.
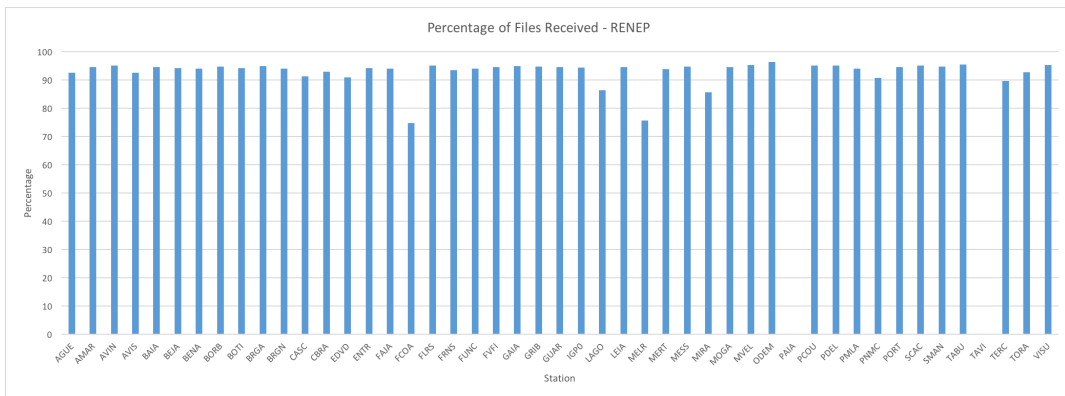
Figure 6.4: downloaded files per station (ReNEP network)

For the SERVIR network, dates used were from 1/10/2015 until 2/04/2016. The reason is the IGeoE has being carried out some internal changes at their system with some outages periods. Therefore, it was decided to not use the data after 02/04/2016 (4376 files). By the time of the writing of this document, the network is back to normal operation mode. Figure 6.5 shows that this network is also very reliable in terms of data. The average percentage of received files is of 79,8% (including 2 stations that were completely off during the time of the test).
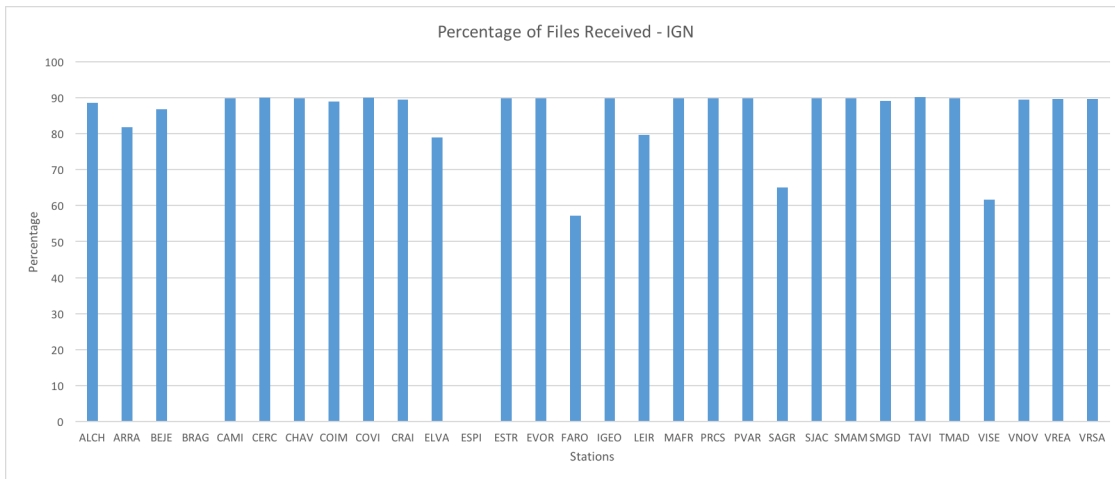
Figure 6.5: downloaded files per station (SERVIR network)

6.1.2 Average delay

Other test performed in terms of network reliability was to assess the average delay (in minutes) taken to receive the files. To calculate this delay is necessary to have two factors in account:

- since there is a two hour delay on the solutions (caused by the orbits/clock availability) the files are marked for download for the first time always with an hour delay (e.g., the file containing the data from 14:00 to 15:00 is collected at 16:00 - this temporal delay is a configurable parameter in the developed scripts);

- based on the factor above, and that the necessary orbits/clock are available only after another hour, the ideal delay in this test would be below the 60 minutes barrier (when the orbit/clock is made available).

The delay represents the difference of the desired date/hour for arrival, and the actual date/hour of arrival. In Appendix B is shown an example of the log that holds this data.

Below it will be presented two figures for every network:

1. The first figure represents the average delay in minutes for every station and the average of the network which is represented by a solid orange line;

2. Since the statistics presented in the first figure can be influenced badly if for some reason a network is off-line for a day or two (e.g. maintenance), it was decided to also compute four statistics:

    - percentage of files arrived without delay;

    - percentage of files arrived with a delay inferior to 15 minutes;

    - percentage of files arrived with a delay inferior to 60 minutes;

    - percentage of files arrived with a delay higher than 60 minutes.

56
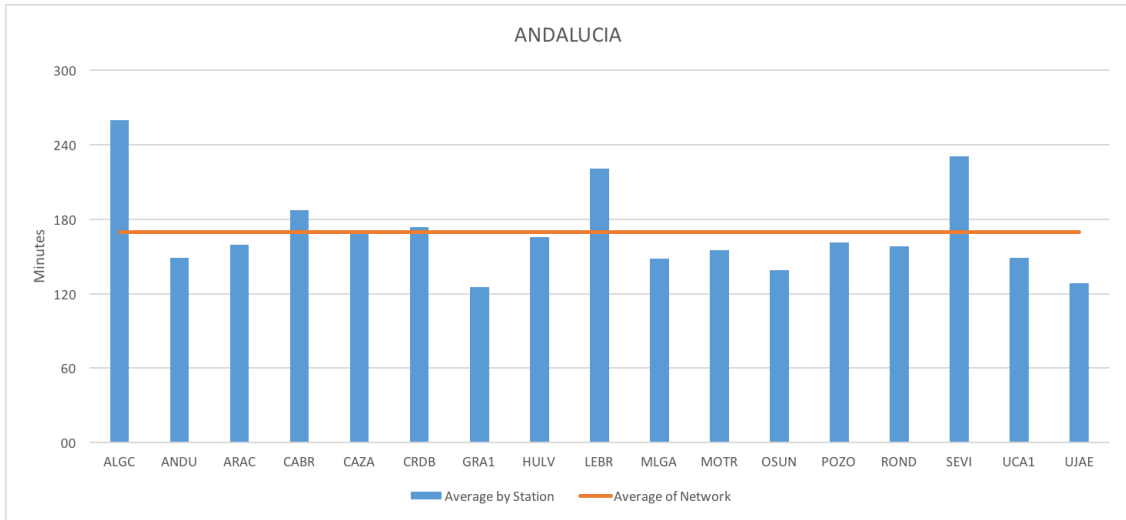
Figure 6.6: Average delay in minutes per station (ANDALUCIA network)

The Andalucia network is once again the worst in our test. The average delay of the network is of approximately 170 minutes (see Figure 6.6).
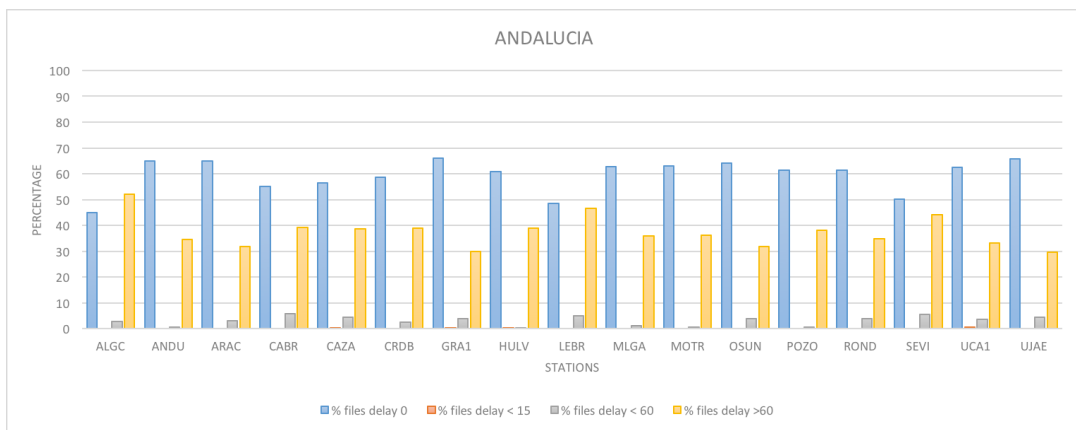
Figure 6.7: Chart with statistics for network Andalucia

Figure 6.7 shows that only 60% of the files arrived during the optimal period for ANDALUCIA.
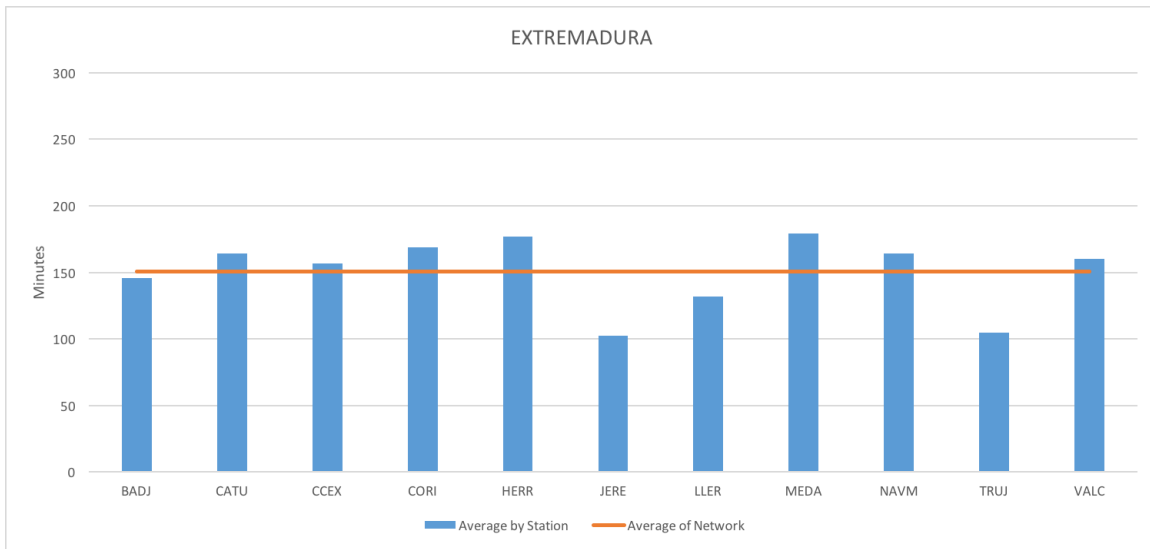
Figure 6.8: Average delay in minutes per station (EXTREMADURA network)

The Extremadura network also performs bad in this test, with an average delay of 150 minutes, as shown in Figure 6.8.
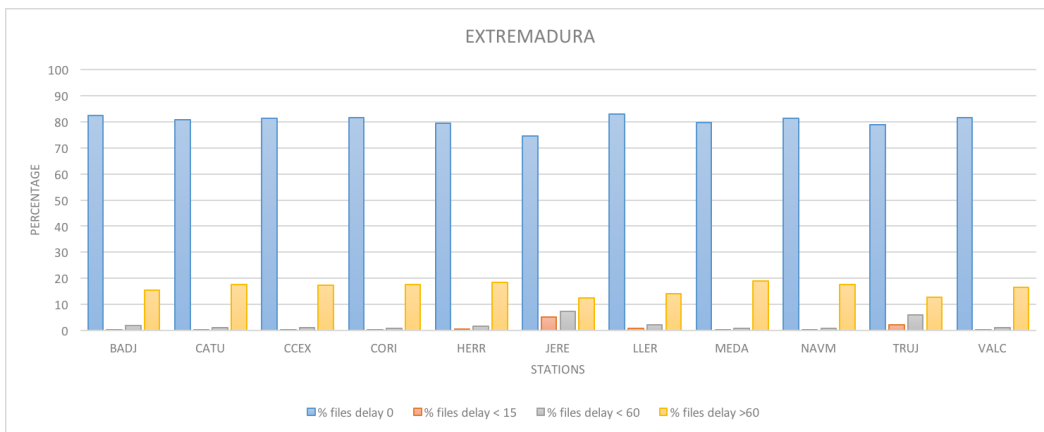
Figure 6.9: Chart with statistics for network Extremadura

Figure 6.9 average delay of the files is mainly caused by station being off-line and not to slow internet connections. In fact around 80% of all the files for this network arrive without delay in normal situations.
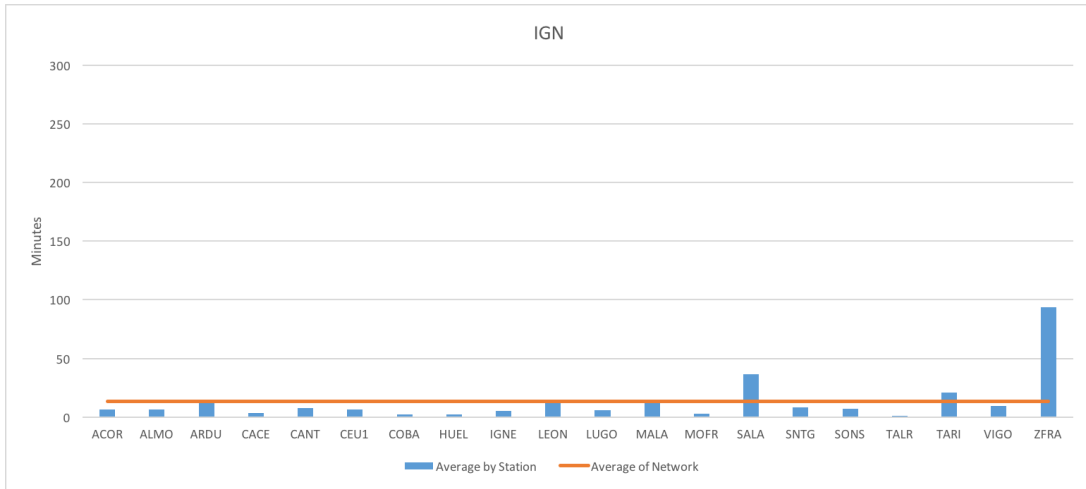
Figure 6.10: Average delay in minutes per station (IGN network)

The IGN network has the best results, with almost all the stations with an average delay below 10 minutess. The average delay of the network is of 13 minutes, as shown in Figure 6.10.
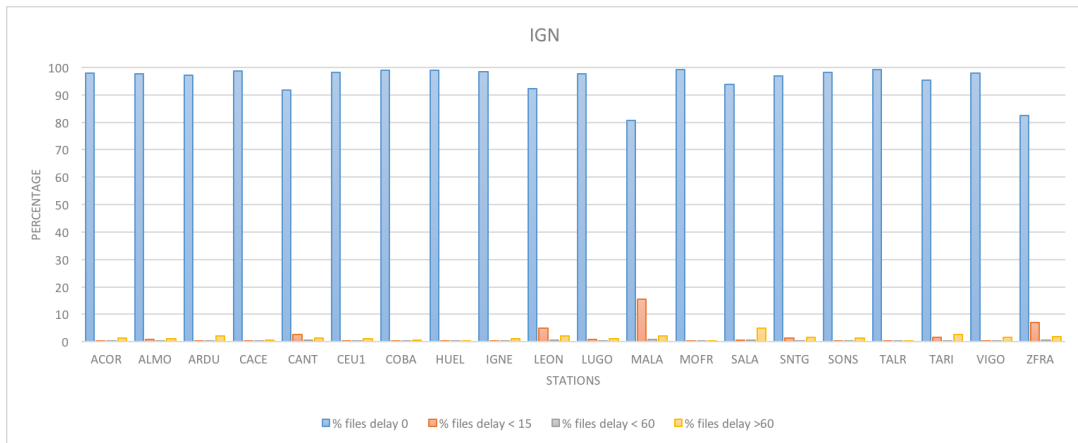
Figure 6.11: Chart with statistics for network IGN

Figure 6.11 show that almost all (above 90% in almost all the stations) the files for this network arrive without delay.
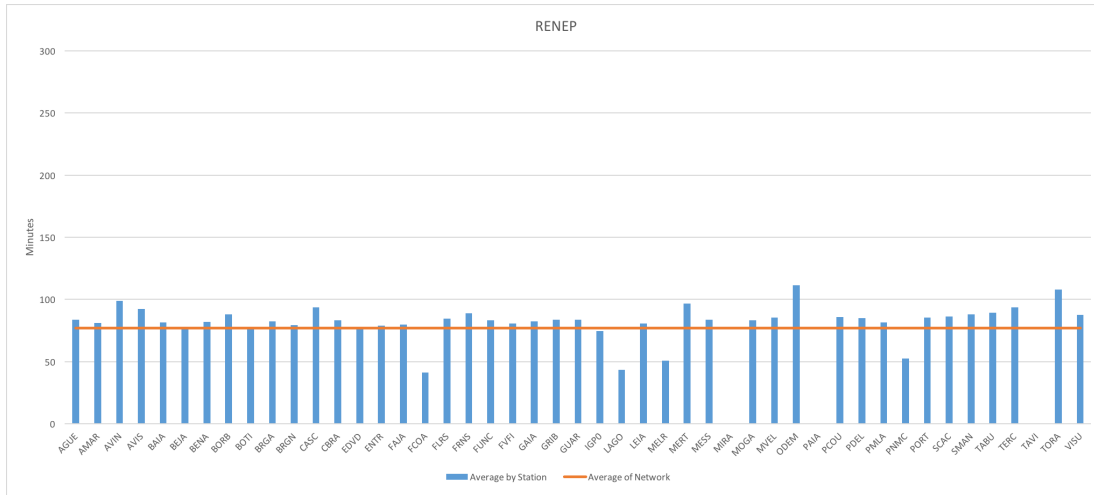
Figure 6.12: Average delay in minutes per station (ReNEP network)

With an average delay of approximately 77 minutes, as shown in Figure 6.12 the ReNEP network provides data in usable time. Even though the average delay is higher than the 60 minutes barrier. Being a large network, sometimes they stop their services for a few hours for maintenance, which influences negatively the average delay.
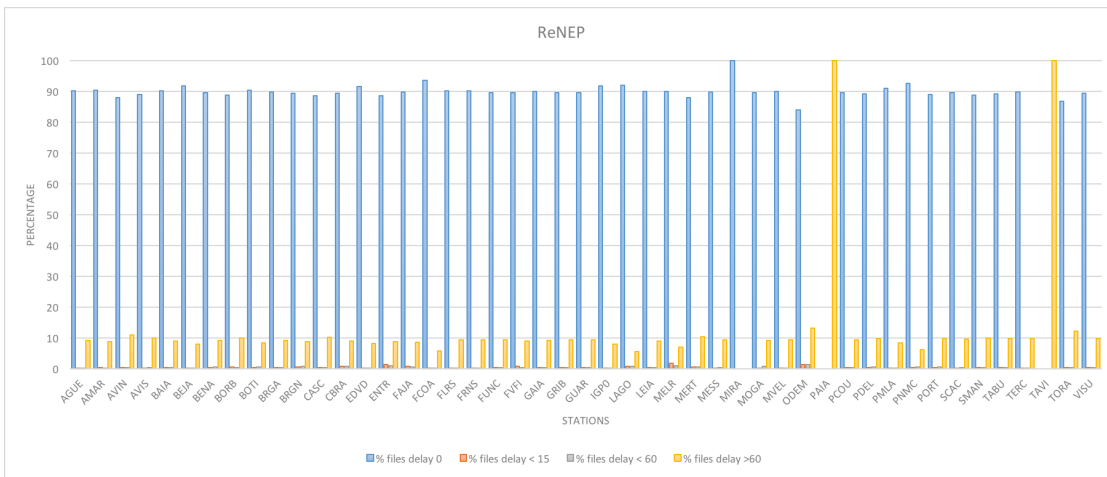
Figure 6.13: Chart with statistics for network ReNEP

Like referred above the average delay was influenced by the maintenance hours. Figure 6.13 illustrates such issue, with about 90% of the files arriving without delay.
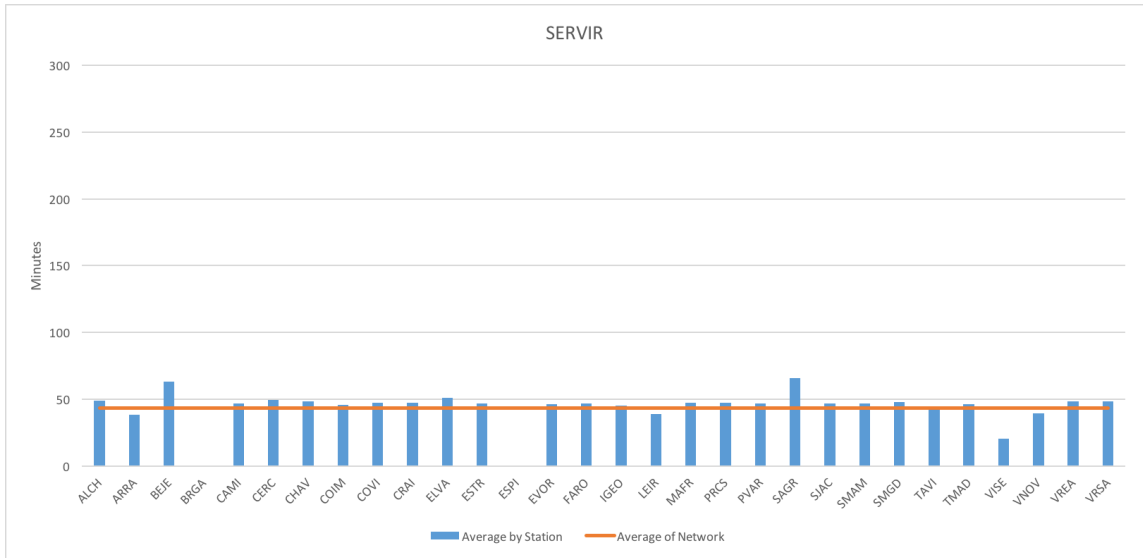
Figure 6.14: Average delay in minutes per station (SERVIR network)

The SERVIR network has an average delay of approximately 44 minutes, as shown in Figure 6.14.
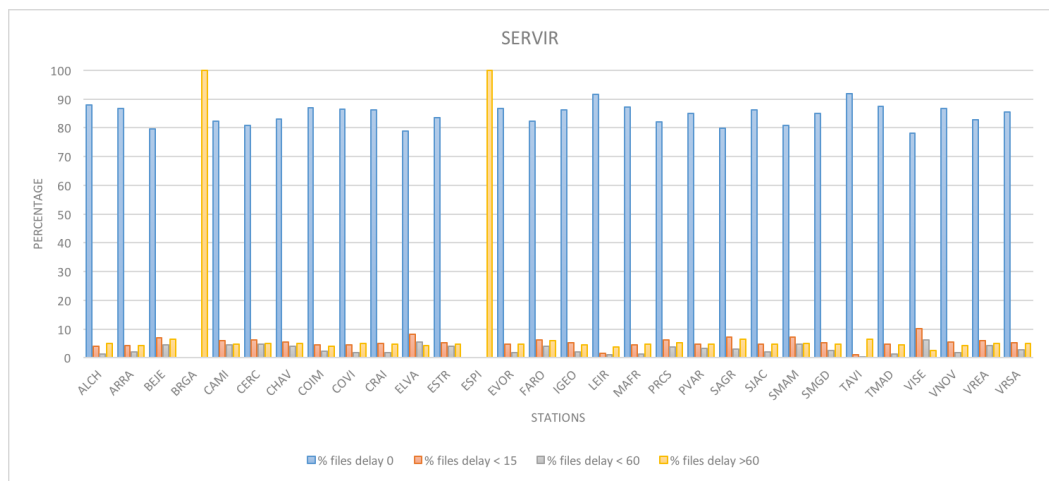
Figure 6.15: Chart with statistics for network SERVIR

Figure 6.15 show that around 80% of all the files arrive without delay.

With these statistics, it can be concluded that these networks combined, make a reliable data source for the project.

6.2   Time Window Influence testing

Longer time series with GNSS observations result in more accurate ZTD estimates. In our case, the limit is set to 30h, which corresponds to the time-span of the ultra-rapid orbits provided by JPL [jpl]. Since hourly files are being used, it is necessary to combine the hourly files for the last 30h into one single 30-hour file. However it might occur that not all files are available. Therefore, one question that someone can ask is:

- "what is the minimum number of solutions that needs to be used to compute a reliable estimate for the last hour?"

It is important to notice that the optimal definition of such threshold value is a complex problem that goes beyond the scope of this M.Sc. project. In fact, it involves the combination of several parameters, that can influence the reliability of the estimates, like:

1. total number of hourly files;

2. distribution of the gaps: concentrated or randomly distributed;

3. percentage of gaps in the last hours;

4. variability of the estimated ZTD/PWV (periods with large variability vs. more stable periods);

5. quality of the observables - how much problems in the observations affect the final solutions;

6. etc..

Therefore, this section presents only the first assessment carried out to obtain a threshold value for the minimal numbers of files used to compute a solution.

The initial idea was to base these tests in the PWV estimates. But in fact in most cases around Europe GNSS-PWV is not used, instead the GNSS-ZTD products are used for assimilation in Numerical Weather Prediction models (like already described in the case of Meteo Office - UK). So even though IPMA wants also the PWV solutions (in order to directly compare with other products), we have done these tests with ZTD in order to to limit this analysis to the GNSS processing only. In [ATB12] it is defined that the ZTD accuracy requirements for nowcasting is between 3 and 10 mm.

6.2.1 Processing with 20 hours

In order to evaluate the influence of less data, only 20h were concatenated to compute the last hour. For that it were chosen 2 stations (randomly) and a set of 5 sequential days.

Figure 6.16: Difference for processing ZTD with 20 or 30 hours for the AGUE station - 05/05/2016 to 09/05/2016

Date/Time	ZTD - 20h	ZTD - 30h	Diff
09/05/16 14:34	2391,9	2390,7	1,2
09/05/16 14:39	2392,2	2391	1,2
09/05/16 14:44	2392,3	2391,1	1,2
09/05/16 14:49	2392,4	2391,2	1,2
09/05/16 14:54	2392,6	2391,4	1,2
09/05/16 14:59	2394,6	2392,6	2
09/05/16 15:04	2395	2393,1	1,9
09/05/16 15:09	2395,7	2393,7	2
09/05/16 15:14	2396,5	2394,5	2
09/05/16 15:19	2397,3	2395,3	2
09/05/16 15:24	2398	2396,1	1,9
09/05/16 15:29	2398,4	2396,6	1,8
09/05/16 15:34	2398,5	2396,7	1,8
09/05/16 15:39	2398,5	2396,8	1,7
09/05/16 15:44	2398,3	2396,6	1,7
09/05/16 15:49	2397,9	2396,2	1,7
09/05/16 15:54	2397,5	2395,8	1,7
09/05/16 15:59	2398,8	2397,6	1,2
09/05/16 16:04	2398,4	2397,2	1,2
09/05/16 16:09	2397,9	2396,6	1,3
09/05/16 16:14	2397,7	2396,4	1,3
09/05/16 16:19	2397,6	2396,3	1,3
09/05/16 16:24	2397,7	2396,3	1,4
09/05/16 16:29	2397,7	2396,3	1,4
09/05/16 16:34	2397,7	2396,3	1,4

Table 6.1: Difference for processing ZTD with 20 or 30 hours for the AGUE station - 05/05/2016 to 09/05/2016

Figure 6.16 is the representation of the results obtained (table 6.1 shows part of the results). As it can be observed the maximum deviation obtained for this station is of 2mm.
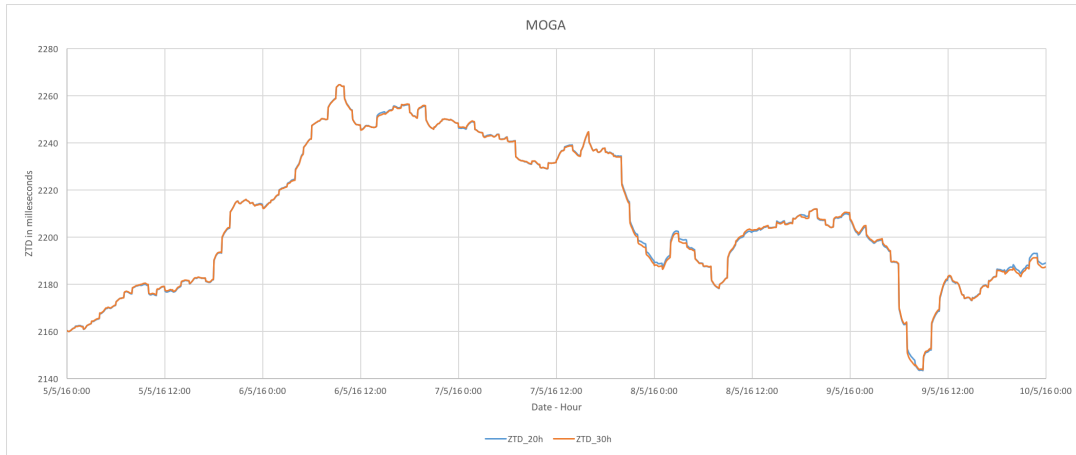
Figure 6.17: Difference for processing ZTD with 20 or 30 hours for the MOGA station - 05/05/2016 to 09/05/2016

Date/Time	ZTD - 20h	ZTD - 30h	Diff
09/05/16 06:49	2163,2	2163,6	0,4
09/05/16 06:54	2163,5	2163,9	0,4
09/05/16 06:59	2152,9	2151,1	1,8
09/05/16 07:04	2151,9	2150	1,9
09/05/16 07:09	2151,1	2149,1	2
09/05/16 07:14	2150,6	2148,5	2,1
09/05/16 07:19	2150,2	2148	2,2
09/05/16 07:24	2149,7	2147,6	2,1
09/05/16 07:29	2149,3	2147,2	2,1
09/05/16 07:34	2149	2146,9	2,1
09/05/16 07:39	2148,7	2146,5	2,2
09/05/16 07:44	2148,3	2146,1	2,2
09/05/16 07:49	2148,1	2145,8	2,3
09/05/16 07:54	2147,7	2145,4	2,3
09/05/16 07:59	2145,6	2145,8	0,2
09/05/16 08:04	2145,1	2145,4	0,3
09/05/16 08:09	2144,7	2145	0,3
09/05/16 08:14	2144,3	2144,6	0,3
09/05/16 08:19	2143,9	2144,2	0,3
09/05/16 08:24	2143,6	2143,9	0,3
09/05/16 08:29	2143,5	2143,8	0,3
09/05/16 08:34	2143,6	2143,9	0,3
09/05/16 08:39	2143,6	2143,9	0,3
09/05/16 08:44	2143,6	2143,9	0,3
09/05/16 08:49	2143,6	2143,9	0,3

Table 6.2: Difference for processing ZTD with 20 or 30 hours for the MOGA station - 05/05/2016 to 09/05/2016

Figure 6.17 is the representation of the results obtained (table 6.2 shows part of the results). As it can be observed the maximum deviation obtained for this station is of 2,3mm.

64

6.2.2 Processing with 10 hours

Since the maximum limit decided (3mm) was not reached in the previous test it was decided to concatenate only 10h and compute the last hour. Another 2, randomly selected, stations were used during a set of 5 sequential days.
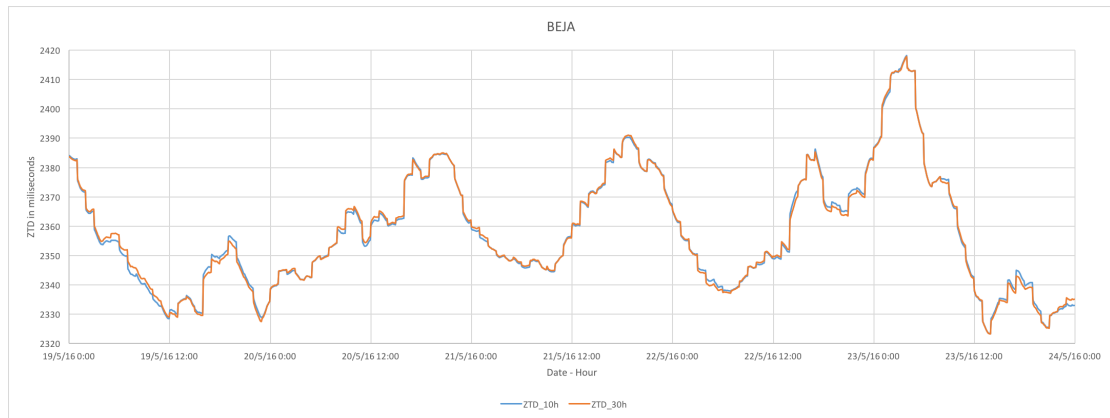
Figure 6.18: Difference for processing ZTD with 10 or 30 hours for the BEJA station - 19/05/2016 to 23/05/2016

Date/Time	ZTD - 10h	ZTD - 30h	Diff
19/05/16 06:40	2349,9	2351,9	2
19/05/16 06:45	2349,8	2351,9	2,1
19/05/16 06:50	2349,9	2351,9	2
19/05/16 06:55	2349,9	2352	2,1
19/05/16 07:00	2345,5	2348,1	2,6
19/05/16 07:05	2345,1	2347,6	2,5
19/05/16 07:10	2344,6	2347,1	2,5
19/05/16 07:15	2344,1	2346,7	2,6
19/05/16 07:20	2343,7	2346,2	2,5
19/05/16 07:25	2343,6	2346,3	2,7
19/05/16 07:30	2343,6	2346,3	2,7
19/05/16 07:35	2343,4	2346,1	2,7
19/05/16 07:40	2343,2	2345,9	2,7
19/05/16 07:45	2343,1	2345,9	2,8
19/05/16 07:50	2343	2345,8	2,8
19/05/16 07:55	2343	2345,8	2,8
19/05/16 08:00	2343,7	2345,5	1,8
19/05/16 08:05	2343,3	2345,1	1,8
19/05/16 08:10	2342,8	2344,6	1,8
19/05/16 08:15	2342,1	2343,8	1,7
19/05/16 08:20	2341,7	2343,4	1,7
19/05/16 08:24	2341,3	2343	1,7
19/05/16 08:29	2340,9	2342,5	1,6
19/05/16 08:34	2340,5	2342,2	1,7
19/05/16 08:39	2340,4	2342,1	1,7

Table 6.3: Difference for processing ZTD with 10 or 30 hours for the BEJA station - 19/05/2016 to 23/05/2016

Figure 6.18 is the representation of the results obtained (table 6.3 shows part of the results). As it can be observed the maximum deviation obtained for this station is of 2,8mm. Which is not a

big increase from the maximum obtained of processing using 20 hour file, and is still below the 3mm barrier.

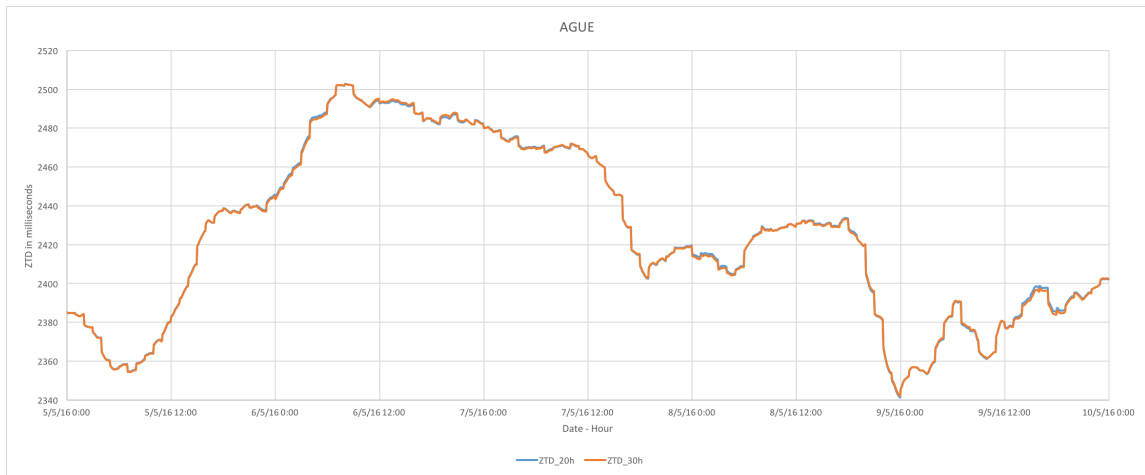
Figure 6.19: Difference for processing ZTD with 10 or 30 hours for the BRGN station - 19/05/2016 to 23/05/2016

Date/Time	ZTD - 10h	ZTD - 30h	Diff
22/05/16 15:54	2247,2	2243,4	3,8
22/05/16 15:59	2250,6	2245,6	5
22/05/16 16:04	2250,2	2245,1	5,1
22/05/16 16:09	2249,8	2244,5	5,3
22/05/16 16:14	2249,6	2244,2	5,4
22/05/16 16:19	2249,4	2243,8	5,6
22/05/16 16:24	2249,1	2243,4	5,7
22/05/16 16:29	2249	2243,1	5,9
22/05/16 16:34	2248,7	2242,7	6
22/05/16 16:39	2248,1	2242,1	6
22/05/16 16:44	2247,7	2241,6	6,1
22/05/16 16:49	2247,2	2241,1	6,1
22/05/16 16:54	2246,9	2240,8	6,1
22/05/16 16:59	2239,8	2234,3	5,5
22/05/16 17:04	2239,3	2233,8	5,5
22/05/16 17:09	2238,9	2233,5	5,4
22/05/16 17:14	2238,5	2233,1	5,4
22/05/16 17:19	2238,2	2232,9	5,3
22/05/16 17:24	2238	2232,7	5,3
22/05/16 17:29	2237,9	2232,6	5,3
22/05/16 17:34	2237,9	2232,5	5,4
22/05/16 17:39	2238,1	2232,7	5,4
22/05/16 17:44	2238,4	2232,9	5,5
22/05/16 17:49	2238,7	2233,2	5,5
22/05/16 17:54	2239	2233,5	5,5

Table 6.4: Difference for processing ZTD with 10 or 30 hours for the BRGN station - 19/05/2016 to 23/05/2016

As it can be observed in Figure 6.19 and table 6.4 the maximum deviation was of 6,1 mm. It can also be observed in Figure 6.19 that such value is not isolated in a single period of time (peak). It shows several periods where the 10h-30h solutions disagree for more than 3mm (minimum threshold value when the variability becomes suspicious of a bad solution). However, it does

not exceed the maximum value suggested by [ATB12].

6.3   Rule definition

Based on the previous tests, it was decided therefore to establish one additional rule: the estimate for the last hour is only done when the last 10h are available in the system.

As previous mentioned, this is not conclusive but it permits to have some minimal guarantees that the estimated solutions computed in operational mode are sufficient reliable.

6.4   GNSS-PWV Quality Testing

The true value of the estimated parameters are not known (otherwise, there was not need to compute them). However, in this particular case, we can use the solutions computed using precise products as reference (i.e., the value that are more probability to be closer to the true). The precise products (orbits and clocks) are normally available 10 to 15 days later at the JPL ftp site [jpl].

These values were used to compare the real-time estimates on a final test. The solutions of 55 RENEP and SERVIR stations were estimated for the entire month of December 2015 (the stations of the ReNEP network were only available until 23/12/2015 due to network maintenance.

The root mean square (r.m.s.) of the daily differences between the estimates of the GNSS-PWV solution in near real time and the GNSS-PWV processed with precise orbits and clocks (that is assumed to provide the correct PWV solutions) was the first statistic calculated.

Figure 6.20 shows the differences for only 5 randomly selected stations for sake of clarity (since the graphic for all 55 stations depicted in Figure 6.21 is too full). The r.m.s. of the differences is below 1,4mm for the entire month (and almost never surpasses 1mm). This is an important fact, since it meets the accuracy requirements of weather nowcasting suggested by [HK06].
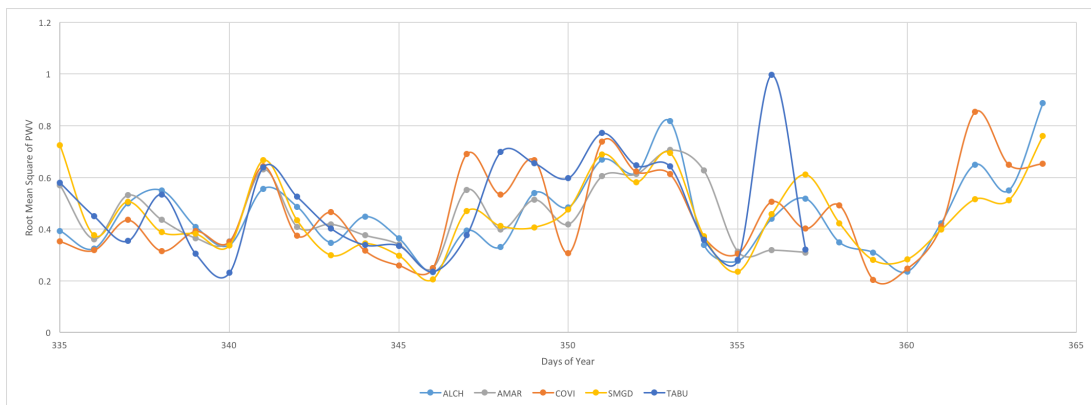
Figure 6.20: Graph with RMS for five stations

Figure 6.21: Graph with RMS of all stations

Another analyzed statistic was the percentage of observations per day with a difference larger than x mm, namely:

1. >= 1 mm of difference;

2. >= 2 mm of difference;

3. >= 3 mm of difference;

SITE	Day of Year Interval	% Diff >= 1	% Diff >= 2	% Diff >= 3
AGUE	335-357	5,59	0,42	0,00
ALCH	335-364	4,60	0,26	0,08
AMAR	335-357	3,93	0,06	0,00
ARRA	335-364	6,48	0,51	0,03
AVIS	335-357	5,40	0,22	0,00
BAIA	335-357	6,41	0,76	0,07
BEJA	335-357	6,87	0,46	0,04
BEJE	335-357	3,81	0,11	0,00
BENA	335-357	2,65	0,21	0,05
BORB	335-357	3,88	0,08	0,00
BOTI	335-357	5,01	0,71	0,18
BRGN	335-357	3,36	0,09	0,00
CASC	335-357	5,44	0,57	0,27
CBRA	335-357	5,51	0,91	0,27
CERC	335-364	7,69	0,44	0,10
CHAV	335-364	4,76	0,28	0,00
COIM	335-364	5,37	0,33	0,03
COVI	335-364	4,82	0,47	0,04
CRAI	335-364	5,51	0,66	0,12
ELVA	335-364	2,93	0,10	0,00
ESTR	335-364	4,41	0,11	0,00
EVOR	335-364	3,70	0,23	0,04
FAJA	335-357	4,42	0,64	0,00
FARO	335-357	3,34	0,16	0,00
FCOA	335-357	4,22	0,15	0,00
FVFI	335-357	5,64	0,22	0,02
GAIA	335-357	7,24	1,11	0,15
GRIB	335-357	4,67	0,14	0,00
GUAR	335-357	2,79	0,17	0,00
IGEO	335-364	11,08	0,48	0,02
IGP0	335-357	5,61	0,27	0,05
LAGO	335-357	3,35	0,11	0,03
LEIA	335-357	5,93	0,85	0,26
LEIR	335-364	6,12	0,64	0,06
MAFR	335-364	5,12	0,27	0,06
MELR	335-357	3,57	0,27	0,00
MERT	335-357	3,15	0,16	0,00

MESS	335-357	3,92	0,18	0,05
MIRA	335-357	6,31	0,36	0,02
MOGA	335-357	4,05	0,02	0,00
ODEM	335-357	4,14	0,15	0,06
PCOU	335-357	7,27	0,39	0,02
PNMC	335-357	3,73	0,19	0,00
PORT	335-357	3,78	0,02	0,00
PRCS	335-364	4,64	0,31	0,03
PVAR	335-364	6,71	0,92	0,15
SAGR	335-364	5,23	0,11	0,00
SCAC	335-357	4,23	0,30	0,14
SJAC	335-364	9,59	1,33	0,18
SMAM	335-364	5,17	0,42	0,17
SMAN	335-357	4,64	0,23	0,03
SMGD	335-364	4,80	0,22	0,01
TABU	335-357	6,09	0,55	0,05
VNOV	335-364	4,54	0,17	0,00
VREA	335-364	8,98	0,81	0,16

Table 6.5: Results of the differences greater than 1mm, 2mm or 3mm in percentage combined by station

Day of Year	% Diff >= 1	% Diff >= 2	% Diff >= 3
335	7,74	0,56	0,14
336	1,84	0,01	0,00
337	4,19	0,25	0,01
338	1,75	0,16	0,08
339	0,33	0,00	0,00
340	1,06	0,00	0,00
341	11,23	0,98	0,19
342	4,65	0,37	0,10
343	0,67	0,00	0,00
344	1,35	0,09	0,01
345	0,47	0,07	0,04
346	0,38	0,00	0,00
347	4,94	0,60	0,00
348	6,57	0,63	0,09
349	7,57	0,36	0,01
350	1,98	0,03	0,01
351	13,92	0,52	0,11
352	11,88	0,81	0,11
353	18,30	1,83	0,24
354	3,00	0,08	0,00
355	0,58	0,00	0,00
356	9,42	0,90	0,18
357	4,59	0,22	0,01

358	5,02	0,31	0,05
359	0,96	0,00	0,00
360	0,44	0,00	0,00
361	3,48	0,19	0,00
362	14,58	1,98	0,17
363	6,15	0,04	0,00
364	14,86	1,17	0,13

Table 6.6: Results of the differences greater than 1mm, 2mm or 3mm in percentage combined by day

The percentage of daily solutions that diverge from reference (the one computed using precise products) is quite small when considering the 1mm treshold (5,31%) and almost neglegilbe when consider 2mm and 3mm treshold values (0,39% and 0,0,6%, respectively). A more detailed analyses per station (Tables 6.5) and per day (Table 6.6) permits to conclude that the variations are larger per day, maybe related to the quality of the JPL products, then by station.

Therefore, it is possible to conclude that the near real-time permits to obtain sufficient accurate (using the precise solutions as reference) estimatives to be used by the meteorologists at IPMA.

6.5   Summary

In this chapter, it were presented some initial tests that permitted to define some of the operational parameters and also to conduct some evaluation of the reliability of the data acquisition and the quality of the computed solutions.

Chapter 7

Conclusions

7.1 Implemented objectives

The main objectives for this dissertation were to implement a near real time service to compute hourly GNSS-PWV solutions, and to study the reliability of the developed system. These objectives were completed with success and the service is currently in operation.

During the course of this dissertation the following scripts were developed to implement the processing data flow:

- Download GNSS data from several providers as soon as that data are available;

- Download Ultra Rapid Orbits and Clocks from JPL;

- Convert the Pressure and Temperature files to to RINEX meteorological format;

- Create RINEX observational files with 30h to be processed by GIPSY;

- Process the GNSS data in order to obtain the ZTD estimates;

- Convert ZTD to PWV;

- Upload the results to IPMA;

- Generate animation with PWV result for the last hour.

On top of these there were also developed some more scripts to help managing system:

- Set a station as off-line if it hasn't provide data to the system in the last 40 hours;

- Check if a station that is off-line is available again;

- Clean the control log files(that holds the missing files to download) to keep them updated with the last 48 hours.

It was developed a dedicated website where the public can see the latest generated animated map of PWV for the Portuguese territory(see http://nuvem.di.ubi.pt). This site also contains general and updated information about the NUVEM project.

It was also implemented a virtualized infrastructure at SEGAL. This new infrastructure will help SEGAL to provide a better support, not just to this project, but also in any other projects where SEGAL's computational resources are necessary.

7.2   System assessment

In addition to the implementation of the system, assessment tests were conducted to establish some main parameters (necessary for the normal operation of the system) and to evaluate its performance, namely:

Network Reliability: It is important to know how to access how reliable the data transference are in order to be able to estimate the tropospheric parameters timely. So, several tests have been carried out to study how many files are missing (in usable time) and what is the medium delay to obtain the desired data. The conclusion is that, despite some data gaps, more significant in some networks (e.g., ANDALUCIA), the processed network at each hour has enough stations to provide a good coverage for the Portuguese territory.

Time window Influence: It is recommended to use GNSS data files (RINEX) of 30 hours (since this is the same time window included in the Orbits and Clock files). However since we are dealing with hourly files, sometimes that is not possible. The analysis of the performed tests (using 10h and 20h files) suggests that a minimal length of 10h should be used to compute the last hour solution.

Quality of GNSS-PWV solutions obtained in near real time: Comparing the near real time results with the post processing solutions, showed that they have similar quality. In fact, most of the solutions have almost no difference. Only occasionally (1%), differences larger than 3mm occur.

7.3   Final notes

It is also important to notice that although the service created is robust, it is still necessary to improve some aspects, namely to estimate the solutions with short delays (<1h) in order to be more effective to be used for nowcasting at IPMA. This would imply to use another software to process the data since it is not possible to do it using the products currently made available by JPL. Even so this product can help IPMA, since the ZTD solutions can be assimilated as one more input in the Numerical Weather Prediction models.

This dissertation also provide a good feedback if in the future JPL would provide a faster Orbit Clock product, or if other product uses a different Software / Orbit, Clock combination to achieve the same goal.

As a personal note, this project was both challenging and rewarding, since the objectives were fulfilled. Not just that the experience acquired during the project will certainly improve my skills, and facilitate my life in the future. I consider my self lucky for having the opportunity to contribute in such an important project of public service that will provide the public with better meteorological data.

7.4 Future Work

Like in any other project there is always room for improvement. In this specific case there are two improvements that would greatly improve the current service:

1. The first is to move the SEGAL public server (geodac) functions (the Pressure and Temperature upload by IPMA, and the uploads made by the networks SERVIR and EXTREMADURA), to the implemented virtualized infrastructure. This will remove the problem existent now that all the data uploaded to SEGAL is entirely dependent on a single physical machine;

2. The second improvement would be the development of a web management platform to interact with the system, and perform tasks like add or remove a station. Not just that this platform could generate automatic statistics, so the administrators can find and debug problems easier, and also generate alerts (e.g. send an email if an entire network has no data for the last N hours).

It would also be important to make a some more tests, mostly in the time window influence, including with PWV estimates.

Bibliography

[AN87]   J. Askne and H. Nordius. Estimation of tropospheric delay for microwaves from surface weather data. *Radio Science*, 22(3):379--386, 05 1987. 7, 8

[ATB12]   Furqan Ahmed, Norman Teferle, and Richard Bingley. An evaluation of real - time, near real - time and post - processed zenith total delay estimates. IGS Workshop 2012, 07 2012. Available from: `https://orbilu.uni.lu/bitstream/10993/14988/1/IGSWS2012_Ahmed.pdf`. 62, 67

[BAD05]   Joël Van Baelen, Jean-Pierre Aubagnac, and Alain Dabas. Comparison of near–real time estimates of integrated water vapor derived with gps, radiosondes, and microwave radiometer. *Journal of Atmospheric and Oceanic Technology*, 22(2):201--210, 02 2005. 5, 6

[BBC+94] Michael Bevis, Steven Businger, Steven Chiswell, Thomas A. Herring, Richard A. Anthes, Christian Rocken, and Randolph H. Ware. Gps meteorology: Mapping zenith wet delays onto precipitable water. *Journal of Applied Meteorology*, 33(3):379--386, 03 1994. 8

[BBH+92] Michael Bevis, Steven Businger, Thomas A. Herring, Christian Rocken, Richard A. Anthes, and Randolph H. Ware. Gps meteorology: Remote sensing of atmospheric water vapor using the global positioning system. *Journal of Geophysical Research*, 97(D14):15787, 1992. 5, 8

[Ben13]   Fábio Bento. Parallelization and implementation of methods for image reconstruction applications in gnss water vapor tomography, 10 2013. 8

[ber]   Bernese gnss software. Available from: `http://www.bernese.unibe.ch/`. 83

[BGA16]   Mike Bowler, Marc Guillemot, and Ahmed Ashour. Htmlunit – welcome to htmlunit, 2016. Available from: `http://htmlunit.sourceforge.net/`. 37

[BNTS06] J. Boehm, A. Niell, P. Tregoning, and H. Schuh. Global mapping function (gmf): A new empirical mapping function based on numerical weather model data. *Geophysical Research Letters*, 33(7), 2006. 7

[BWS06]   Johannes Boehm, Birgit Werl, and Harald Schuh. Troposphere mapping functions for gps and very long baseline interferometry from european centre for medium-range weather forecasts operational analysis data. *Journal of Geophysical Research: Solid Earth*, 111(B2):n/a--n/a, 02 2006. 7

[con11]   Converting hosts, 09 2011. Available from: `https://docs.oracle.com/cd/E20065_01/doc.30/e18549/ptov.htm`. 25

[COS]   COST. Cost. Available from: `http://www.cost.eu/`. 17

[cre11]   Creating a server pool, 2011. Available from: `https://docs.oracle.com/cd/E27300_01/E27309/html/vmusg-svrpool-create.html`. 25

[exp11]   Exploitation of ground-based gps for climate and numerical weather prediction applications, 05 2011. Available from: `http://www.cost.eu/COST_Actions/essem/716`. 17

[gam13]   Gamit-globk homepage, 2013. Available from: `http://www-gpsg.mit.edu//gtgk/docs.htm`. 85

[Gao06]   Yang Gao. Precise point positioning and its challenges, aided-gnss and signal tracking. *Inside GNSS*, pages 16--18, 11 2006. Available from: `http://www.insidegnss.com/auto/NovDec06GNSSSolutions.pdf`. 9, 10

[gip]   Gipsy-oasis. Available from: `https://gipsy-oasis.jpl.nasa.gov/`. 11

[gps]   gpstk. gpstk. Available from: `http://www.gpstk.org/bin/view/Documentation/WebHome`. 82

[gps11]   Gps.gov: signals, 2011. Available from: `http://www.gps.gov/systems/gps/modernization/civilsignals/`. 5

[Gro01]   The PHP Group. Php: Releases, 2001. Available from: `https://secure.php.net/releases/`. 31

[Gue03]   G. Guerova. Application of gps derived water vapour for numerical weather prediction in switzerland, 2003. 6, 7

[Hed09]   Jonathan Hedley. Jsoup java html parser, with best of dom, css, and jquery, 2009. Available from: `https://jsoup.org/`. 37

[HK06]   S De Haan and Netherlands KNMI. National/regional operational procedures of gps water vapour networks and agreed international procedures. Technical report, 2006. Available from: `https://www.wmo.int/pages/prog/www/IMOP/publications/IOM-92_GPS-WaterVapour/IOM-92_GPS-Water-Vapour.pdf`. 67

[igs]   Igs orbits. Available from: `https://igscb.jpl.nasa.gov/components/prods_cb.html`. 11, 12

[ipm15a] Ipma: Arome, 10 2015. Available from: `https://www.ipma.pt/pt/enciclopedia/otempo/previsao.numerica/index.html?page=arome.xml`. 13

[ipm15b] Ipma: Ecmwf, 10 2015. Available from: `https://www.ipma.pt/pt/enciclopedia/otempo/previsao.numerica/index.html?page=ecmwf.xml`. 13

[ipm15c] Ipma: Nwp, 10 2015. Available from: `https://www.ipma.pt/pt/enciclopedia/otempo/previsao.numerica/index.html`. 15

[ipm15d] Ipma: Radiosondagem, 10 2015. Available from: `https://www.ipma.pt/pt/educativa/observar.tempo/index.jsp?page=altitude.xml`. 13, 14

[ipm15e] Ipma: Satellite images, 10 2015. Available from: `https://www.ipma.pt/pt/enciclopedia/otempo/observacao.remota/index.html`. 14

[Jon10]   Jonathan Jones. An assessment of the quality of gps water vapour estimates and their use in operational meteorology and climate monitoring, 05 2010. 18

[jpl]   Jpl orbits. Available from: `https://gipsy-oasis.jpl.nasa.gov/index.php?page=data`. 11, 12, 61, 67

[Kap96]   E.D. Kaplan. *Understanding GPS-principles and <pplications.* Boston: Artech House, 1996. 5

[KIA+12] Morten A Køltzow, Karl-Ivar Ivarsson, Solfrid Agersten, Lars Meuller, Dag Bjørge, Ole Vignes, Bjart Eriksen, Per Dahlgren, Martin Ridal, and Rebecca Rudsar. Metcoop meteorological co-operation on operational nwp (numerical weather prediction). *METCOOP MEMO*, (01), 2012. Available from: `http://metcoop.org/memo/2012/01-2012-METCOOP-MEMO.PDF`. 13

[LSB+13] K. Lagler, M. Schindelegger, J. Böhm, H. Krásná, and T. Nilsson. Gpt2: Empirical slant delay model for radio space geodetic techniques. *Geophysical research letters*, 40(6):1069--1073, 03 2013. 7

[nes16] Nesdis blended total precipitable water (tpw) products, 05 2016. Available from: `http://gis.ncdc.noaa.gov/geoportal/catalog/search/resource/details.page?id=gov.noaa.ncdc:C00803`. 16

[Nie96] A E Niell. Global mapping functions for the atmosphere delay at radio wavelengths. 101(2):3227--3246, 1996. Available from: `http://www.haystack.mit.edu/geo/pubs/NMF_JGR.pdf`. 7

[Nie01] A.E. Niell. Preliminary evaluation of atmospheric mapping functions based on numerical weather models. *Physics and Chemistry of the Earth, Part A: Solid Earth and Geodesy*, 26(s 6–8):475–480, 2001. Available from: `http://www.sciencedirect.com/science/article/pii/S1464189501000874`. 7

[oqu15] O que é o java?, 05 2015. Available from: `http://www.java.com/pt_BR/about/whatis_java.jsp`. 37

[pol15] Policy-based cluster and capacity management, 07 2015. Available from: `https://docs.oracle.com/database/121/CWADD/pbmgmt.htm#CWADD92894`. 22

[RHH05] C Rizos, M B Higgins, and S Hewitson. New gnss developments and their impact on survey service providers and surveyors. 2005. Available from: `https://www.fig.net/organisation/comm/5/gnss/rizos_etal2005a.pdf`. 5

[rtk] Rtklib: An open source program package for gnss positioning. Available from: `http://www.rtklib.com/`. 81

[rtn] Rtnet. Available from: `http://www.gps-solutions.com/rtnet_software`. 85

[Saa72] J Saastamoinen. Atmospheric correction for troposphere and stratosphere in radio ranging of sate. *The use of artificial satellites for geodesy*, 15:247, 1972. Available from: `http://adsabs.harvard.edu/abs/1972GMS....15..247S`. 7

[SD14] Jan Van Sickle and John Dutton. Differencing, 2014. Available from: `https://www.e-education.psu.edu/geog862/node/1727`. 9

[TBHS07] Volker Tesmer, Johannes Boehm, Robert Heinkelmann, and Harald Schuh. Effect of different tropospheric mapping functions on the trf, crf and position time-series estimated from vlbi. *Journal of Geodesy*, 81(6-8):409--421, 01 2007. 7

[UNA16a] UNAVCO. Teqc, 2016. Available from: `https://www.unavco.org/software/data-processing/teqc/teqc.html`. 37

[UNA16b] UNAVCO. Unavco, 04 2016. Available from: `https://www.unavco.org/`. 37

Appendix A

GNSS Processing Software

In this appendix a detailed description on the several GNSS Processing Software available is made. These descriptions were taken from the respective webpages (as referenced).

A.1 RTKLIB

RTKLIB is an open source program package for standard and precise positioning with GNSS (global navigation satellite system). RTKLIB consists of a portable program library and several APs (application programs) utilising the library. The features of RTKLIB are:

It supports standard and precise positioning algorithms with: GPS, GLONASS, Galileo, QZSS, BeiDou and SBAS;

It supports various positioning modes with GNSS for both real-time- and post-processing: Single, DGPS/DGNSS, Kinematic, Static, Moving-Baseline, Fixed, PPP-Kinematic, PPP-Static and PPP-Fixed;

It supports many standard formats and protocols for GNSS: RINEX 2.10, 2.11, 2.12 OBS/NAV/ GNAV/HNAV/LNAV/QNAV, RINEX 3.00, 3.01, 3.02 OBS/NAV, RINEX 3.02 CLK, RTCM ver.2.3, RTCM ver.3.1, RTCM ver.3.2, BINEX, NTRIP 1.0, NMEA 0183, SP3-c, ANTEX 1.4, IONEX 1.0, NGS PCV and EMS 2.0;

It supports several GNSS receivers□ proprietary messages: NovAtel, Hemisphere, u-blox , Sky-Traq, JAVAD, Furuno and NVS;

It supports external communication via: Serial, TCP/IP, NTRIP, local log file (record and play-back) and FTP/HTTP (automatic download);

It provides many library functions and APIs (application program interfaces): Satellite and navigation system functions, matrix and vector functions, time and string functions, coordinates transformation, input and output functions, debug trace functions, platform dependent functions, positioning models, atmosphere models, antenna models, earth tides models, geoid models, datum transformation, RINEX functions, ephemeris and clock functions, precise ephemeris and clock functions, receiver raw data functions, RTCM functions, solution functions, Google Earth [34] KML converter, SBAS functions, options functions, stream data input and output functions, integer ambiguity resolution, standard positioning, precise positioning, post-processing positioning, stream server functions, RTK server functions, downloader functions.

RTKLIB main developer is Takasu Tomoji (T. Takasu) from the Laboratory of Satellite Navigation, Tokyo University of Marine Science and Technology.[rtk]

A.2 GPSTk

GPSTk project aims to provide an open source library and suite of applications to the satellite navigation community--to free researchers to focus on research, not lower level coding.

GPS users employ practically every computational architecture and operating system. Therefore the design of the GPSTk suite is as platform-independent as possible. Platform independence is achieved through use of the ISO-standard C++ programming language. The principles of object-oriented programming are used throughout the GPSTk code base in order to ensure that the code is modular, extensible and maintainable.

The GPSTk suite consists of a core library, auxiliary libraries, and a set of applications. The GPSTk provides a wide array of functions that solve processing problems associated with GNSS such as processing or using standard formats such as RINEX. The libraries are the basis for the more advanced applications distributed as part of the GPSTk suite.

The features of GPSTk:

RINEX utilities The RINEX utilities provide a set of applications that can be used to examine, manipulate, and plot RINEX observation files;

Positioning The positioning applications include two different applications that perform standard pseudorange-based positioning and two that implement differential phase-based solutions;

Residual analysis A residual analysis application computes two types of measurement residuals using a single receiver or two receivers in a zero baseline configuration.

Ionospheric modeling The ionospheric modelling applications utilize the two frequency TEC estimate from the RINEX utilities and compute a model of the ionosphere.

Signal Tracking Simulation These utilities simulate the tracking of GPS C/A and P-code.

Basic transformations Conversions of time and coordinate systems.

Observation data collection and conversion Translating receiver specific data formats to RINEX.

File comparison and validation Differing observations files against a truth source.

Data editing Simple editing like systematic removal of observations by satellite, type or time and more advanced editing like cycle slip detection and correction.

Autonomous and relative positioning Navigation and surveying applications.

The GPSTk is sponsored by Space and Geophysics Laboratory, within the Applied Research Laboratories at the University of Texas at Austin (ARL:UT). GPSTk is the by-product of GPS research conducted at ARL:UT since before the first satellite launched in 1978; it is the combined effort of many software engineers and scientists. In 2003, the research staff at ARL:UT decided to open source much of their basic GNSS processing software as the GPSTk. [gps]

A.3 BERNESE

The Bernese GNSS Software is a scientific, high-precision, multi-GNSS data processing software developed at the Astronomical Institute of the University of Bern. It is, e.g., used by CODE (Center for Orbit Determination in Europe) for its international (IGS) and European (EUREF/EPN) activities.

Features[ber]:

Processes all static GNSS applications;

RINEX (2.x and 3.x) data interface;

Processes all principal observables recorded by high precision geodetic receivers, i.e., code and phase data on both carriers;

Takes into account differential code biases (DCBs) for satellites and receivers for applications involving code observations;

Single and dual frequency data may be processed in the same estimation step, use of ionosphere models to minimize the impact of ionospheric biases on station coordinates;

Five different linear combinations of L1 and L2 may be used;

Allows for a fully combined processing of data from GPS, GLONASS, and GPS/GLONASS receivers;

The use of observations for Galileo, SBAS, Compass, and QZSS is prepared;

Different ambiguity resolution strategies allow fixing of phase ambiguities on up to very long (several thousand kilometers) baselines – available for GPS and GLONASS;

Processing in double-difference mode as well as in zero-difference mode. Particularly suitable for time transfer and precise point positioning;

Processing and combination of data from various receiver types in the same processing step;

Kinematic and reduced-dynamic precise orbit determination for Low Earth orbiters (LEOs);

Capability to process SLR observations to GNSS and LEO satellites;

Station motions modeled for plate motions, tides, and ocean loading.

Geophysical deformation models can be introduced as grids and validated by estimating scaling factors;

Different troposphere mapping functions available, estimation of troposphere gradient parameters;

Higher order ionosphere corrections (2nd and 3rd order and ray bending) including the estimation of scaling factors;

Modeling of all time-dependent parameters by a piece-wise linear, continuous representation, in particular for troposphere zenith delay and gradient parameters, Earth orientation parameters, and global ionosphere models;

Allows applying and estimating antenna phase patterns for receiver (GNSS-specific) and for satellite antennas. Supports the ANTEX antenna information exchange format;

The parameter estimation programs may be used for baseline/session/campaign/multiple-campaign processing. Many different complex solutions using (e.g., daily) normal equation systems can be produced without reprocessing the observations;

The normal equation stacker (ADDNEQ2) offers sophisticated options for pre-elimination of parameters, parameter transformations, and for the definition of the geodetic datum, including minimum constraint network conditions;

Setup and manipulation of normal equations without inversion;

Sophisticated observation simulation tool available;

The Bernese Processing Engine (BPE) allows to tailor automated processing sequences. It includes options supporting, in particular, the efficient reprocessing of GNSS data from several years

A.4   RTNet

This software is described by having two main components:

1. real-time -epoch-server (RTES) which reads an epoch of GPS data from GPS receivers, translates, compresses, transmits via the internet to the processing computer, uncompresses and makes data available to the real-time network processing software (RTnet);

2. RTnet takes the observation, and processes the data;

The main characteristics of the RTnet are:

RTnet is designed for processing of GPS networks with the highest possible accuracy. Besides the careful modeling of all effects that may influence the GPS measurements an important feature of the RTnet model is the correct handling of the measurement dependencies (RTnet works with the full variance-covariance matrices and not only with their baseline-related parts);

RTnet is primarily designed for real-time applications (however, post-processing is possible, too). In order to achieve high numerical stability of the estimation process a square-root filter is used instead of a simple Kalman filter;

RTnet can estimate a large number of parameter types. In addition to the ambiguities and station coordinates the list of parameter types includes the receiver and satellite clock corrections, parameters that characterize the ionospheric and tropospheric biases etc;

TRtnet can process data in network mode or in PPP mode in real-time or post processing;

Real-time estimation of satellite clocks for regional or global networks is possible without the requirement of a stable reference clock (the average of all satellite clocks in view from the network is used as reference);

Real-time results from one master Rtnet run (i.e. satellite clocks or tropospheric parameters) can be read directly on input by other slave rtnet runs;

Processing was tested up to 10 Hz data rate;

In addition to its original main purpose (precise network deformation monitoring) RTnet is capable to estimate DGPS corrections and generate virtual reference station data and FKP corrections for the network region;

RTnet processes zero-differenced observations and the satellite and receiver clock corrections are estimated at every epoch independently. Due to one-to-one correlation with the clock biases the phase ambiguities may be resolved on the double-difference level. The double differencing of GPS measurements is optimized every epoch and the information on the resolved ambiguities is stored.

The software can use precise GPS orbits, broadcast orbits, or both combined. For real-time applications this implies that some other process can provide predicted ultra-rapid orbits from the IGS (so called IGU orbits). The software will always use the latest predicted file. If broadcast orbits are used then these will be read from the RTES server. If both orbits shall be combined then the logic works as follows: RTnet will use the IGU orbits for all available satellites, and for those satellites for which the IGU orbits are not available RTnet will use the broadcast orbits.[rtn]

A.5 GAMIT

GAMIT/GLOBK is a comprehensive GPS analysis package developed at MIT, the Harvard-Smithsonian Center for Astrophysics (CfA), Scripps Institution of Oceanography (SIO), and Australian National University for estimating station coordinates and velocities, stochastic or functional representations of post-seismic deformation, atmospheric delays, satellite orbits, and Earth orientation parameters. Although the software is currently maintained by the three authors of this document at MIT, many people have made substantial contributions. The orbital integration and modules used in computing the theoretical phase observable have their origins in the Planetary Ephemeris Program (PEP) written by Michael Ash, Irwin Shapiro, and Bill Smith at Lincoln Laboratory in the 1960's, with later contributions by Bob Reasenberg and John Chandler at MIT. The codes for processing GPS observations were developed at MIT in the 1980's by Chuck Counselman, Sergei Gourevitch, Yehuda Bock, Rick Abbot, and King. GAMIT attained its current form through the efforts of Bock, Danan Dong, Peng Fang (SIO), Kurt Feigl, Herring, King, McClusky (ANU), Mike Moore (ANU), Peter Morgan (Canberra U), Mark Murray (NM Tech), Liz Petrie (U Newcastle), Berkhard Schraffin (Ohio State), Seiichi Shimada (NEID), Paul Tregoning (ANU), and Chris Watson (U Hobart). GLOBK was developed by Herring at CfA for combination of VLBI data and modified at MIT to incorporate GPS data. Details of these contributions may be found in the references listed at the end of this manual. Funding for the early development of GAMIT was provided by the Air Force Geophysics Laboratory, and for GLOBK by NASA. Current support for development and support of the scientific community comes primarily from the National Science Foundation through UNAVCO.[gam13]

Appendix B

Files and Logs of the System

B.1 NETWORK files

```
ague,Agueda,AGUE
amar,Amareleja,AMAR
avin,ArrudaVinhos,AVIN
avis,Avis,AVIS
baia,Baiao,BAIA
beja,Beja,BEJA
bena,Benafim,BENA
borb,Borba,BORB
boti,Boticas,BOTI
brag,Braga,BRGA    1 2 3
brgn,Braganca,BRGN
casc,Cascais,CASC
cbra,CBranco,CBRA
edvd,EDVSJMadeira,EDVD
entr,Entroncamento,ENTR
faja,Fajao,FAJA
fcoa,VNFozCoa,FCOA
flrs,Flores,FLRS
frns,Furnas,FRNS
func,Funchal,FUNC
fvfi,FValeFigueira,FVFI
gaia,VNGaia,GAIA
grib,GloRibatejo,GRIB
guar,Guarda,GUAR
igp0,IGP,IGP0
```

Figure B.1: stations_network file - ReNEP

Figure B.1 shows part of the file that contains the stations info for the ReNEP network, where:

1. Represents the 4 char code of the station for the network;

2. Represents the name of the station;

3. Represents the 4 char code at SEGAL.

In this case it is necessary 3 fields for each station. The reason is that the 4 char code (that should be unique but is not, because there isn't an world organism responsible for this) for some station are already at use in SEGAL (SEGAL collects data all over the world), so an internal re-name is necessary.

```
acor,ACOR
lugo,LUGO
vigo,VIGO
sntg,SNTG
leon,LEON
ardu,ARDU
cant,CANT
igne,IGNE
sons,SONS
almo,ALMO
coba,COBA
mala,MALA
ceu1,CEU1
tari,TARI
mofr,MOFR
talr,TALR
sala,SALA
cace,CACE
zfra,ZFRA
huel,HUEL
```

Figure B.2: Stations_network file - IGN

Figure B.2 shows part of the file that contains the station of the IGN network, as shown it only need 2 fields, name and 4 char code.

As mentioned in Chapter 5 there is a stations_network_off file for each network. The composition of this files has the same rules as the ones shown in the previous figures.

B.2   Log for files missing

```
ftp://ftp.igeo.pt/Beja/rinex/2015/09/02/beja245x.zip
ftp://ftp.igeo.pt/Mirandela/rinex/2015/09/02/mira245x.zip
ftp://ftp.igeo.pt/Mogadouro/rinex/2015/09/02/moga245x.zip
ftp://ftp.igeo.pt/MontemorVelho/rinex/2015/09/02/mvel245x.zip
ftp://ftp.igeo.pt/Odemira/rinex/2015/09/02/odem245x.zip
ftp://ftp.igeo.pt/Palmela/rinex/2015/09/02/palm245x.zip
ftp://ftp.igeo.pt/PCoura/rinex/2015/09/02/pcou245x.zip
ftp://ftp.igeo.pt/PDelgada/rinex/2015/09/02/pdel245x.zip
ftp://ftp.igeo.pt/Penamacor/rinex/2015/09/02/pena245x.zip
ftp://ftp.igeo.pt/Portalegre/rinex/2015/09/02/port245x.zip
ftp://ftp.igeo.pt/SantCacem/rinex/2015/09/02/scac245x.zip
ftp://ftp.igeo.pt/SaoMancos/rinex/2015/09/02/sman245x.zip
ftp://ftp.igeo.pt/Tabuaco/rinex/2015/09/02/tabu245x.zip
ftp://ftp.igeo.pt/Tavira/rinex/2015/09/02/tavi245x.zip
ftp://ftp.igeo.pt/Terceira/rinex/2015/09/02/terc245x.zip
ftp://ftp.igeo.pt/Torrao/rinex/2015/09/02/torr245x.zip
ftp://ftp.igeo.pt/Viseu/rinex/2015/09/02/vise245x.zip
```

Figure B.3: downloads_missingNETWORK log file - ReNEP

Figure B.3 shows the example of a log file for missing files (that the system tries to download. As it can be observed each line represents a missing file, so the system reads the line (that is the file url) and tries to download it.

B.3   Log for statistics

```
ague276r.zip/18:01/18:01/276/276
ague276s.zip/19:01/19:01/276/276
ague276t.zip/20:01/20:01/276/276
ague276u.zip/21:01/21:01/276/276
ague276v.zip/22:01/22:01/276/276
ague276w.zip/23:01/23:01/276/276
ague276x.zip/00:01/00:01/277/277
ague277a.zip/01:01/01:01/277/277
ague277b.zip/02:01/02:01/277/277
ague277c.zip/03:01/03:01/277/277
ague277d.zip/04:01/04:01/277/277
ague277e.zip/05:01/05:01/277/277
ague277f.zip/06:01/06:01/277/277
ague277g.zip/07:01/07:01/277/277
ague277h.zip/08:01/08:01/277/277
ague277i.zip/09:01/09:01/277/277
ague277j.zip/10:01/10:01/277/277
ague277k.zip/11:01/11:01/277/277
ague277l.zip/12:01/12:01/277/277
ague277m.zip/13:01/13:01/277/277          1 2 3 4 5
```

Figure B.4: statistics log file - station AGUE

Figure B.4 shows the example of a statistics logs for a station where:

1. Represents the name of the file downloaded;

2. Represents the hour it was expected to be downloaded;

3. Represents the hour it arrived;

4. Represents the day of the year it was expected to be downloaded;

5. Represents the day of the year it arrived;

As it can be observed each line holds info for a file.