TOR VERGATA UNIVERSITY

Geoinformation PhD Programme

XXIII Cycle

Atmospheric probabilistic retrieval through GNSS signals

Andrea Antonini

October 2013

Mentor:
Prof. Giovanni Schiavon

Co-Mentor:
Dott. Alberto Ortolani

Coordinator:
Prof. Domenico Solimini

"...Theory is when we know everything but nothing works.
Praxis is when everything works but we do not know why.
We always end up by combining theory with praxis:
nothing works and we do not know why..."

unknown author but frequently attributed to Albert Einstein

To My Family

Contents

List of Figures

List of Tables

List of Abbreviations

3D-Var: Three Dimension Variational

4D-Var: Four Dimension Variational

APL: Applied Physics Laboratory

ARP: Antenna Reference Point

ART: Algebraic Reconstruction Techniques

A-S: Anti-Spoof

AUTONAV: AUTOnomous NAVigation

BPSK: Binary Phase Shift Keying

CDMA: Code Division Multiple Access

CDDIS: Crustal Dynamics Data Information System

COSMEMOS: COoperative Satellite navigation for MEteo-marine MOdelling and Services

CS: Control Segment

CS: Commercial Service

DCB: Differential Code Bias

DD: Double Difference

DGPS: Differential GPS

DWD: Deutscher WetterDienst

ECEF: Earth Centred Earth Fixed

ECI: Earth Centred Inertial

ECMWF: European Centre for Medium Range Weather Forecasts

EMP: ElectroMagnetic Pulse

ESA: European Space Agency

EU: European Union

EUREF: EUropean REference Frame

FDMA: Frequency Division Multiple Access

FOC: Full Operational Capability

FP7: 7^{th} Framework Programme

GA: Ground Antenna

GBD: Global Burst Detector

GEOS-5 DAS: Goddard Earth Observing System Model, Version 5 Data Assimilation Scheme

GLONASS: GLObal NAvigation Satellite System

GMAO: Global Modeling and Assimilation Office

GMF: Global Mapping Function

GMT: Greenwich Mean Time

GNSS: Global Navigation Satellite System

GPS : Global Positioning System

GSA: European GNSS Agency formerly European GNSS Supervisory Authority

GSI: Grid-point Statistical Interpolation

GTRF: Galileo Terrestrial Reference Frame

HF: High Frequencies

HP: High Precision

ICD: Interface Control Ducument

IEEE: Institute for Electrical and Electronic Engineers

IERS: International Earth Rotation and Reference System Service

IGS: International GNSS Service

IMF: Isobaric Mapping Functions

IOV: In Orbit Validation

IPP: Ionospheric Pierce Point

IPW: Integrated Precipitable Water

IRI: International Reference Ionosphere

IWV: Integrated Water Vapour

JPL: Jet Propulsion Laboratory

MART: Multiplicative Algebraic Reconstruction Technique

MCS: Master Control Station

MDS: Minimum Detectable Signal

MEO: Medium Earth Orbit

MERRA: Modern Era Retrospective-analysis for Research and Applications

MM5: NCAR Meososcal Model

MS: Monitor Station

MW: MicroWaves

NASA: National Aereonautic and Space Administration

NESDS: NASA's Earth Science Data Systems

NAVSTAR: NAVigation System with Time And Ranging

NAVWAR: NAVigation WARfare

NGS: National Geodetic Survey

NCST: Naval Center for Space Technology

NDS: NUDET Detection System

NNSS: Navy Navigation Satellite System

NMF: Niell Mapping Function

NRL: Naval Research Laboratory

NUDET: NUclear DETonation

NWM: Numerical Weather Models

NWP: Numerical Weather Prediction

OCS: Operational Control Segment

OCXO: Oven Controlled Crystal Oscillator

OS: Open Service

PCV: Phase Centre Variation

PLL: Phase Locked Loop

PPP: Precise Point Positioning

PPS: Precise Positioning Service

PRN: Pseudo Random Noise

PRS: Public Regulated Service

PS: Surface Pressure

PW: Precipitable Water

RAP: Rapid Refresh

RC: Ranging Codes

RH: Relative Humidity

RHCP: Right-Hand Circularly Polarised

RINEX: Receiver Independent Exchange Format

RMSE: Root Mean Square Error

RUC: Rapid Update Cycle

SA: Selective Availability

SD: Single Difference

SAR: Search And Resque

SBAS: Satellite Based Augmentation System

SHD: Slant Hydrostatic Delay

SLP: Sea Level Pressure

SOL: Safety Of Life

SSM: Spread Spectrum Modulation

SP: Standard Precision

SPD: Slant Path Delay

SPS: Standard Position Service

STD: Slant Total Delay

SWD: Slant Water Delay

SWV: Slant Water Vapour

TAI: International Atomic Time

TCXO: Temperature Compensated Crystal Oscillator

TEC: Total Electron Content

TIMATION: TIMe navigATION

TOT: Top Of Troposphere

UHF: Ultra High Frequency

UV: UltraViolet

UTC: Universal Time Coordinated

VLBI: Very Long Baseline Interferometer

VMF1: Vienna Mapping Functions 1

WADGPS: Wide Area Differential GPS

WGS84: World Geodetic System 1984

WV: Water Vapour

WVR: Water Vapour Radiometers

ZD: Zero Difference

ZHD: Zenith Hydrostatic Delay

ZTD: Zenith Tropospheric Delay

ZPD: Zenith Path Delay

ZWD: Zenith Wet Delay

Acknowledgements

I would like to thank Prof. Giovanni Schiavon who supported me professionally in these years and the Geoinformation programme PhD coordinator Prof. Domenico Solimini for having given the opportunity to carry out this research work within the TorVergata University.
My thanks go to Dr. Alberto Ortolani, who had the basic idea of this research work, and with determination encouraged and guided during this period me and his research team. His contribution to this work has been priceless both on a technical and personal level, for which I am extremely grateful. Many thanks to Dr. Riccardo Benedetti for his professionalism and expertise, very useful for the understanding of many aspects of this work. I am grateful to Dr. Rovai who has given an essential contribution in the implementation of some data processing procedures of the present work, and to Dr. Stefano Romanelli who contributed via his skills by the set up of the database used for data collection and management for this study. I thank Dr. Samantha Melani for the advices and especially for her precious help in various stages of scripts debugging. I would like to thank the LAMMA Consortium and its staff, that have invested in this research activity making possible the development of this work. Many thanks to the people who in the course of this time contribute to clarify some technical aspects: Prof. Kay Borre, Prof. Mauro Leonardi, Dr. Gunnar Elgered and Dr. Mauro Grossi.

Finally but primarily, thanks a lot to my wife Elisabetta and my son Luca who have patiently endured the time I have taken from them, during the development of this work.

This work has been partially developed within the COoperative Satellite navigation for MEteo-marine MOdelling and Services (COSMEMOS) project which also has founded part of its research activities. COSMEMOS is a 7^{th} Framework Programme (FP7) project, managed by the European GNSS Agency (GSA), finalised to improve the safety of maritime navigation, through the development of technology and services, based on the European satellite navigation system Galileo.

Introduction

The Global Navigation Satellite System (GNSS) receivers are used in many applications. May be surprising to learn that among these applications there is also aerology. In fact aerology is the term indicating the study of the air and the atmosphere system by means of measurements coming from instruments like balloons, sounders, radiometers, radars, lidars, ground-based meteorological observations, and other systems. It is finalised on the study of the vertical status of the atmosphere, not limited to the stratus nearest the earth surface. This research work stays into this context through the use of ground-based GNSS receivers, a type of instruments whose development has not been specifically designed for the study of the atmosphere and weather phenomena. The interaction of the GNSS satellites signals and the atmosphere, both ionosphere and troposphere was an aspect widely studied since the beginning of the development of the Global Positioning System (GPS): the first satellite navigation system. These aspects were well known because they critically impact on the positioning techniques. Indeed, the localisation of a receiver over the Earth's surface is based on the principle of triangulation, i.e. the time and space coordinates are computed by using the travel time of signals coming from at least four satellites, for which the position and the signal transmission time are known. All the sources that introduce a delay on the satellite receiver travel time, directly affect the positioning accuracy. Many research activities have demonstrated the feasibility of the calculation of these delays, with related increasing positioning accuracies. In parallel, the result of some research work (the first was [1]) has shown that the signal delay caused by water vapour in troposphere can be estimated and used in combination with coincident (in time and space) ground-based temperature and pressure measurements, to retrieve the total column water vapour, i.e. the Integrated Precipitable Water (IPW). As the water vapour has a fundamental role both in the hydrological cycle, by triggering the weather phenomena, and in the greenhouse effect, by trapping long wavelength radiation emitted by the Earth's surface, the knowledge of the temporal behavior of IPW is a huge potential opportunity. The availability of such ground-based network of GNSS receivers actually coincides with the availability of sensors for the estimation of the water content in the atmosphere. Several weather centres have developed operational schemes for the assimilation of GNSS-based water vapour observations in numerical weather prediction

models. The main limitations of this kind of GNSS-IPW retrievals is that they do not provide information about the profile of the vertical distribution of water vapour. For this reason the last frontier of this science is oriented towards the attempt of reconstruction of three-dimensional content of water vapour, commonly through the use of dense networks of GNSS receivers and using tomographic approaches.

In the present research work (whose some theoretical aspects have been addressed in [2] and the results of a preliminary implementation were shown in [3]) a new retrieval approach is presented for the estimation of vertical distribution of meteorological variables. It is based on a Bayesian inference approach, which combines the information content of an extended ten-years dataset with the instantaneous set of measurements for retrieving the desired quantities with an associated probability of occurrence. Practically for the quantities to be estimated a set of possible "states" has been built by considering the output of a reanalysis meteorological model during the period of a decade at synoptic times. The frequencies distribution of this state vectors constitute the prior distribution. For the same period the ground-based measurements of pressure, temperature and humidity and the GNSS-based Zenith Tropospheric Delay (ZTD) estimation, have been collected. The combination of model outputs and measurements leads to the likelihood function, to be used directly in the Bayes theorem to retrieve the posterior distribution, whenever occurs the availability of new measures. The retrieval process assigns a weight, namely the posterior probability distribution, to each possible state vector, and consequently to each profile of atmospheric temperature, pressure and humidity content.

The method has been tested using data from some test sites, over the Mediterranean area, selected on the basis of given requirements, such as the availability of a meteorological station, a GNSS station, and of course, data from the meteorological model, over the ten-year period in order to construct the prior dataset, in addition to the availability of balloon observations for validation purposes. On the basis of these requisites the selected sites have been the coastal sites of Cagliari and Ajaccio, the latter in an area of complex orography, and the inland site of Medicina, which is located in a flat area.

A full description of the technique is provided in this work from a theoretical point of view and then providing the numerical results of an operational-like application over a full year time period. The obtained retrievals are compared against balloon observations, in order to validate the method and to assess the related errors, even if over the few (three) available sites.

The technique is straightfully applicable in anywhere a weather station provides surface measurements of pressure, temperature and humidity and a GNSS receiver gives an estimate of the tropospheric delay. The growing availability of both measures

of meteorological stations and networks of geodetic GNSS stations, increases its applicability, presenting it as a tool with increasing potentialities. As lots of numerical weather models are provided with assimilation scheme able to ingest both IPW/ ZTD observation and atmospheric vertical profiles, they can theoretically ingest also the outputs of this retrieval algorithm. It is therefore hoped that an instrument of this type may contribute to the creation of more and more complete weather patterns for numerical weather prediction models initialisation.

This non conventional type of profiler sensor could be a mean to help the reconstruction and localisation of the stage of the convection development, turbulent mixing, and clouds formation processes in operational weather forecasts. In fact all this information are directly determined by the vertical distribution of water vapour, temperature and pressure. The main advantage comes from the fact that this technique is virtually costless if one consider taking advantage of the high number of GNSS stations network available throughout the Earth's surface for different scopes. In any case, even the installation of new sensors is advantageous from an economic point of view compared to radiosounding stations or ground-based radiometers installations, because of the more consolidated and improved technologies of GNSS receiving stations.

A very short description of the content of each chapter of this work is given below:

- **Chapter 1**: Provides a description of the atmospheric quantities object of the retrieval and the main relationship (used in the work) between them; the mechanisms of interaction of the atmosphere system with MicroWaves (MW) signals, mainly focusing on those of interest for GNSS application are also shown. The subdivision of the atmosphere in ionosphere and troposphere is clarified and their relative effects introduced in the GNSS signal are described.

- **Chapter 2**: Is an introduction to the GNSS, including also a brief history of satellite positioning systems, and giving a description of the functioning principles, the present configuration and the future developments. An analysis of the various error sources is given together with the methods used to compute them finalised to obtain the desired quantity: the tropospheric delay. Finally a very basic algorithm for positioning is described.

- **Chapter 3**: With logical continuity with the previous chapter a description of the method for the computation of the Zenith Path Delay (ZPD) (i.e. ZTD), starting from the GNSS receiver measurements is provided, with reference to the most commonly used software. Main applications of the GNSS Meteorology are listed and briefly discussed to give an overview of the state of the art of this research field.

- **Chapter 4**: The description of the developed retrieval method is given, starting from the theoretical aspects, up to the technical description of the test implementations, and the relative results. The data sources used in this work and the processing strategies motivating the necessary choices are here described. The processing results are shown for some relevant cases.

- **Chapter 5**: In this chapter a quantitative evaluation of the performances of the method is given. The information gain, when the number of observations increases, is explained by means of entropy changes. A comparison between the entropy values relative to different scenarios and conditions is also given. The validation of the retrieved quantity is given by comparison with balloon measurements, providing the values of correlation, Root Mean Square Error (RMSE) and bias. A comparison with the most diffuse GNSS-based method for Water Vapour (WV) content estimation [1] is also provided.

Chapter 1

Interaction of the GNSS signal with the Atmosphere

1.1 The atmosphere and the principal state variables

1.1.1 Classification of atmospheric regions

The atmosphere can be schematised as made by concentric layers which differ depending on physical and chemical characteristics and classified according to such properties of the application into consideration. The first stratification can be made considering the homogeneity of atmospheric composition; the homosphere is the portion of the Earth's atmosphere, up to an altitude of about 90 km above sea level, in which there is continuous turbulent mixing, and hence the composition of the atmosphere is relatively constant; as opposed to the heterosphere, the higher regions in which exists the diffusive equilibrium condition and the composition varies with altitude with increasing presence of rarefied light gases such as helium and hydrogen nearly up to reach the interplanetary vacuum. A classification of the atmospheric layers and their thickness depending on the main applications on which they are involved, is proposed in [4]. Fig. 1.1 shows this classification.

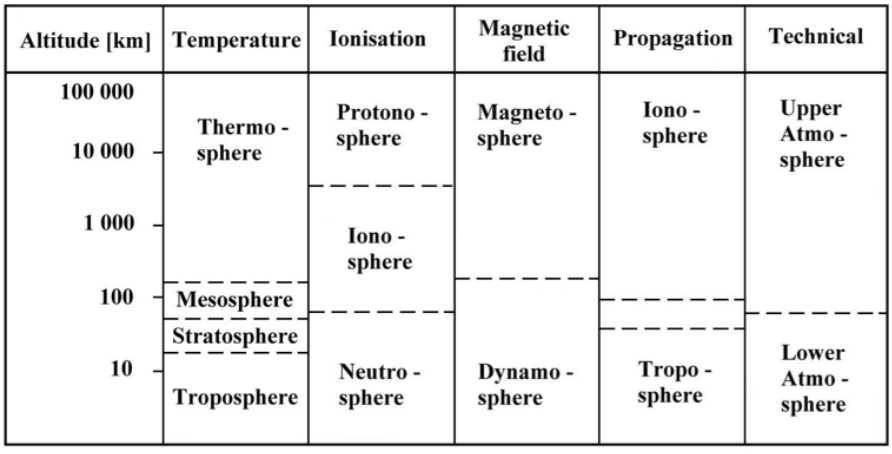


Figure 1.1: Earth atmosphere layers.

In this work the analysis is focused on the impact of the atmosphere on the radiowaves propagation at L band frequencies (1 to 2 GHz), in which GNSS satellite positioning signals are working. However, the purpose of this study is to extract information related to meteorology, from radiowave signals propagating in the atmosphere. Therefore with reference to Fig. 1.1, the applications that we will consider are propagation and temperature, as in meteorology the stratification of the atmosphere essentially comes from the mean vertical profile of temperature.

From the radiowaves propagation point of view, the atmosphere is composed by two main layers: troposphere and ionosphere.

- Troposphere: The troposphere is the lowest region of Earth's atmosphere ranging from 0 to $40 - 50$km. In this layer the charged particles are virtually absent and the molecules are well mixed, so that the troposphere can be considered a neutral gas. As we will see in this layer the refraction index is slightly greater than 1 and is dependent on the value of the atmospheric pressure, temperature and water vapour partial pressure. The troposphere is a non dispersive medium for the GNSS radiowaves, and the refraction index can be considered virtually constant with frequency.

- Ionosphere: The Earth ionosphere is named so because it is slightly ionised by solar radiation. The ionosphere is thus a shell of electrons and electrically charged atoms and molecules that surrounds the Earth, stretching from a height of about 50 km to more than 1000 km.

From a meteorological point of view the atmosphere has a different subdivision, which is mainly determined by the climatological (i.e. time averaged) height gradient of temperature:

- Troposphere: In meteorology the height of the troposphere is up to about 14 - 18 km (depending on latitude with a maximum at the equator and a minimum at poles). It is characterised by a negative temperature gradient in function of height ($\nabla_h T < 0$).

- Tropopause: It is the boundary between the troposphere and the layer above, where occurs the change of sign of the temperature gradient.

- Stratosphere: The stratosphere has an extension ranging from the end of tropopause up to about 50 km. In this region the temperature increases with heigh ($\nabla_h T > 0$). It is more rarefied than troposphere.

- Stratopause: it is the portion of atmosphere where a relative maximum temperature of about 0° C occurs.

- Mesosphere: It extends from the end of stratopause up to about 80 -100 km. Temperature again decreases. It is characterised by a weakly ionised layer which absorbs the cosmic radiation and essentially marks the beginning of the ionosphere.

- Mesopause: it is a separation layer where the temperature reaches the minimum value in the whole atmosphere (about $-100°$ C).

- Thermosphere: It follows the mesopause up to a maximum height of 500 km. The temperature undergoes a strong increase.

- Thermopause: It marks the separation between the portion of atmosphere where the main components are not more heavier gases (like nitrogen and oxygen), but particles with large free path (hydrogen and helium). It is the reference layer for the computation of the solar constant[1]. The height of the thermopause varies between about 200 and 500 km depending on solar activity. It is also called exobase because is the base of the surrounding layer.

- Exosphere: Is the layer between the thermopause and about 10000 km, beyond which begins the interplanetary vacuum. The gas particles that reach and exceed the escape velocity no longer take part in the rotation of the Earth and disperse in space.

The troposphere is the seat of most of meteorological phenomena (clouds, precipitations, lightnings,...) because $\nabla_h T$ can assume values bringing to vertical instability, and contains approximately 75% of the atmosphere's mass and almost all (99%) of its water vapour and aerosols. However it should be point out that sometimes in the stratosphere between 20 and 30 km, it can be observed the presence of iridescent or nacreous clouds. These are sporadic cases that however indicate the presence of water vapour in the stratosphere. In sum about 99% of neutral air mass and all the water vapour are contained in the first 50 km of the atmosphere; thus troposphere and stratosphere have a very similar behavior from the radiowave propagation point of view, for this reason they are commonly classified as a single layer in GNSS theory.

In the following chapters we will refer only to the troposphere and ionosphere. Both the troposphere and the ionosphere refract GNSS satellite signals and are sources of error for GNSS applications. This will be explained in more details below as the subject of the second section of the present chapter.

[1]The solar constant is the amount of radiation that reaches the Earth from the Sun per unit of time and surface (hence it is a power per unit area), measured in a plane perpendicular to the rays. Its mean value is about 1.367 W m^{-2}, varying in function of the sun Earth distance.

1.1.2 Atmospheric quantities and main relationships

The atmosphere is the fluid system which constitutes a casing of the Earth's crust. Mixtures distributions and motions of the gas constituents of the atmosphere are the main responsible of the weather phenomena, that happen in the lower layers where most of atmospheric mass is concentrated (about 90% in the first 17 km and about 99% in the first 50 km). The mean proportion of gas concentrations by volume in the low atmosphere (0 to 50 km) is given in Table 1.1.

Constituent	**Fractional concentration**
Nitrogen (N_2)	78.08%
Oxygen (O_2)	20.95%
Argon (Ar)	0.93%
Water Vapour (H_2O)	$0-5\%$
Carbon dioxide (CO_2)	380 ppm
Neon (Ne)	18 ppm
Helium (He)	5 ppm
Methane (CH_4)	1.75 ppm
Krypton (Kr)	1 ppm
Hydrogen (H_2)	0.5 ppm
Nitrous Oxide (N_2O)	0.3 ppm
Ozone (O_3)	$0-0.1$ ppm

Table 1.1: Selected Merra model files content ([5]).

Besides the gases that occur in very high concentrations (N_2, O_2 and Ar), water vapour plays a very important role in the triggering of weather phenomena. The mass of water vapour is about 0.25% of the atmospheric mass, but it's concentration is highly variable in time and space depending among other variables also on the temperature, that defines the maximum amount of water vapour potentially present (i.e the saturation concentration, see Eq. 1.13), ranging from very low concentrations in the cold regions up to 5% by volume in hot air masses. This variability combined with the fact that the water vapour stores energy in the form of latent heat[2], makes it one of the major responsible of development and dynamics of weather systems.

As the atmosphere is a gas mixture its thermodynamic status can be univocally described by its pressure, density and temperature. The relationship between these three is the equation of state, that changes from one gas to another based on the properties of the molecules that compose it. According to Dalton's law *the total pressure exerted by a mixture of gases that do not interact chemically is equal to the sum of partial pressure of the gases.* For the atmosphere distinction is made between dry air, i.e. the gas mixture which excludes the water vapour, considering the other constituents, and

[2]Evaporation and condensation are processes requiring and releasing heat respectively.

the water vapour, namely wet air. Starting from the ideal gas law,

$$P = \rho \; R \; T \tag{1.1}$$

where P is the pressure, ρ the density, T the temperature and R is the ideal gas constant (universal gas constant $R = 8.3145$ J K^{-1} kg^{-1}). The specific gas behaviour is considered introducing the dependence on the specific type of gas in the constant R. At $T = 0$ ° C for dry air $R_d \simeq 287.053$ J K^{-1} kg^{-1}, and for water vapour $R_w \simeq 461.5$ J K^{-1} kg^{-1}. Both R_d and R_w slightly vary as a function of temperature as better described in § 1.3.1.

A gas in a gravitational field where pressure gradient force balances the gravity force is in hydrostatic equilibrium, that is commonly represented by the hydrostatic equation:

$$dP = - \; \rho \; g \; dz \tag{1.2}$$

Moving vertically from a height z to $z + dz$ the pressure must vary by an amount dP which is linked to the product of the air density ρ and g (taking into account the gravity). The hydrostatic equilibrium condition may be taken as almost always true, except if we are in the presence of strong vertical motions of the atmosphere (as for example in the case of storm). Pressure at a height level z can be derived integrating Eq. 1.2 as:

$$P(z) = - \int_z^\infty dP = \int_z^\infty \rho \; g \; dz \tag{1.3}$$

The pressure at an height z is equal to the weight per unit area of the overlying air in the vertical column.

The geopotential is a quantity given by the work needed to carry a mass of 1 kg from the sea level to a defined height z. It depends only on the height z of that point. Rescaling the geopotential quantity by the globally averaged gravity acceleration at the Earth surface $g_0 = 9.81$ m s^{-2}, the geopotential height is obtained.

$$h = \frac{1}{g_0} \int_0^z g(lat, lon, \zeta) \; d\zeta \tag{1.4}$$

where the gravity acceleration $g(lat, lon, z)$ is expressed in function of latitude lat, longitude lon, and geometrical height z. In meteorology the geopotential height is commonly used instead of the geometrical height and the outputs of models provides the vertical distribution of the physical quantities as a function of geopotential height. Alternatively these quantities are expressed as a function of pressure levels, and in this case also the geopotential height is expressed in function of pressure levels. The advantage of using the geopotential height is the possibility to use an averaged constant value of the gravity acceleration g_0. The conversion from the geopotential height to geometrical height can be done taking into account the law of universal gravitation and

using Eq. 1.4, obtaining the relationship:

$$h = \frac{z\ R_T(lat, lon)}{R_T(lat, lon) + z} \tag{1.5}$$

where the quantity $R_T(lat, lon)$ is the Earth radius at the specific site of latitude lat and longitude lon.

As previously mentioned, in the atmosphere constituents a distinction is made between dry air and water vapour content. Partial pressure values of dry air (P_d) and water vapour (e_w) can be added to give the total atmospheric pressure according to the Dalton's law. Sometimes referring to dry air and using the gas law (Eq. 1.1), a quantity called virtual temperature T_v can be introduced to account for the presence of water vapour, by expressing the total air pressure as:

$$P = \rho\ R_d\ T_v \tag{1.6}$$

and T_v as:

$$T_v = \frac{T}{1 - \frac{e_w}{P}(1 - \epsilon)} \tag{1.7}$$

where ϵ is the quantity given by the ratio between molecular weights of water (M_w) and dry air (M_d):

$$\epsilon = \frac{R_d}{R_w} = \frac{M_w}{M_d} = 0.622 \tag{1.8}$$

The combination of the gas law expressed in the form of Eq. 1.6, the hydrostatic equation (Eq. 1.2) and the geopotential height definition (Eq. 1.4) leads to the hypsometric equation:

$$h_2 - h_1 = \frac{R_d}{g_0} \int_{P_2}^{P_1} T_v \frac{dP}{P} \simeq \frac{R_d}{g_0}\ \overline{T_v}\ \ln \frac{P_1}{P_2} = H \ln \frac{P_1}{P_2} \tag{1.9}$$

where $\overline{T_v}$ is the average virtual temperature between height levels h_1 and h_2 and $H = 29.3\ \overline{T_v}$ is the scaled height. Water vapour is a minor constituent of the atmosphere, but it plays a crucial role in the energy balance and transportation in the atmosphere. In fact part of the energy coming from solar radiation can be stored as latent heat through state transition to water vapour, and released during condensation. Latent heat transfer typically takes place in the convection and advection motion. The amount of water vapour in the atmosphere can be expressed through many different parameters. The presence of water vapour in the atmosphere has been previously treated as vapour pressure e_w and its effect on the air density has been accounted through the virtual temperature (Eq. 1.7).

Sometimes we refer to mixing ratio as the ratio between water vapour mass (m_w) and dry air mass (m_d):

$$r = \frac{m_w}{m_d} = \frac{\epsilon e_w}{P - e_w} \tag{1.10}$$

where ϵ is defined in Eq. 1.8. In the atmosphere the magnitude of r can range from a few grams per kilogram at medium latitude to values of 20 g kg^{-1}.

The ratio between the water vapour mass and the mass of total air is the specific humidity q:

$$q = \frac{m_w}{m_d + m_w} \simeq \frac{\epsilon e_w}{P} \tag{1.11}$$

Typical values of e_w range from zero to a few hectoPascal, and values of P are of order of hundreds of hectoPascal; therefore from Eqs. 1.10 and 1.11 it follows that the values of r and q are nearly equivalent. The absolute humidity ρ_w is the concentration of water vapour in air and is given by the mass of water vapour respect to unit of volume, essentially a partial density:

$$\rho_w = \frac{m_w}{V} \simeq \frac{e_w}{R_w\, T} = \frac{e_w}{P}\, \epsilon\, \rho_d \tag{1.12}$$

Any change of temperature or pressure involves a change in the volume of air, and consequently in the absolute humidity value, not necessarily as a consequence of changes in the content of water vapour.

The concentration of water vapour rarely exceeds the saturation level, a threshold beyond which tends to condense very fast and to reach an equilibrium between condensation and evaporation. The saturation vapour pressure e_s varies with temperature and can be expressed by the Arden Buck semi-empirical relationship:

$$e_s\,(T_c) = 6.1121 \exp\left[\left(18.678 - \frac{T_c}{234.5}\right)\left(\frac{T_c}{257.14 + T_c}\right)\right] \tag{1.13}$$

where T_c is the temperature expressed in Celsius degrees.

When as a result of a temperature reduction the absolute humidity reach the saturation level there is a condensation of the water vapour and the thermal value is the dew point temperature. The dew point temperature is the threshold temperature at which, at constant pressure, the air becomes saturated.

The Relative Humidity (RH) is the ratio between the actual mixing ratio and the saturation mixing ratio expressed as percentage. Some relationship can be extracted from the previous equations:

$$RH = 100\, \frac{r}{r_s} \;= 100\, \frac{e_w}{e_s} \;=\; 100\, \frac{\rho_w}{\rho_s} \;=\; 100\, \frac{q}{q_s} \tag{1.14}$$

The dew point is thus the temperature at which the saturation mixing ratio become equal to the actual mixing ratio, so it holds:

$$RH = 100\, \frac{r_s(T_d, P)}{r_s(T, P)} \tag{1.15}$$

As a consequence, given the measurement of the dew point temperature T_d (in Celsius), a formula to compute the water vapour partial pressure is simply:

$$e_w = 6.1121 \exp\left[\left(18.678 - \frac{T_d}{234.5}\right)\left(\frac{T_d}{257.14 + T_d}\right)\right] \tag{1.16}$$

1.1.3 Profiles of atmospheric variables

The pressure exponential variation can be derived from Eq. 1.9, in function of the geometrical height z that replaces the geopotential height h:

$$P(z) = P_0 \exp\left[-\frac{h}{H}\right] = P_0 \exp\left[-\frac{zR_T}{(R_T+z)H}\right] \tag{1.17}$$

given the pressure measurement at surface level (P_0). Atmospheric pressure decreases with height, due to the fact that the height of overlying air column becomes gradually smaller, and that in the upper layers the atmosphere is more rarefied. Therefore also the vertical profile of air density exponentially decreases as a function of height. The patterns of the vertical profiles of density and pressure are very similar. The previous equations are used by meteorologists to estimate the height of pressure levels of the atmosphere.

The temperature profile has a significantly different pattern compared to the pressure one, with different behavior at different vertical layers, mainly due to the absorption emission characteristics of the sun-Earth-atmosphere system. The atmosphere partially reflects, partially absorbs and partially is transparent to the radiation emitted by the sun. The sun radiation reaching the Earth surface is again absorbed and reflected and finally reemitted (Planck's law) as infrared outgoing longwave radiation.

As previously mentioned the mean vertical profile of temperature is responsible of the division of the atmosphere in five regions: troposphere, stratosphere, mesosphere, thermosphere, exosphere. Between one layer and the contiguous one there are transition layers respectively named: tropopause, stratopause, mesopause, thermopause.

In the troposphere the temperature decreases as a function of height, due to convection of the air because it is mainly heated by the Earth surface emission. Height gradient is limited by the convection phenomenon, when air expands and rises, to be replaced by upper cooler air, which begins the process all over again. The vertical thermal gradient is named lapse rate Γ ($\Gamma = -\frac{dT}{dz}$). Considering an adiabatic process lapse rate (i.e. assuming no heat transfer to or from air parcel), for an air parcel rising adiabatically without condensation, the temperature change is associated to the dry adiabatic lapse rate $\Gamma_d = 9.8$ K km^{-1}. If we take into consideration the atmosphere humidity, as the condensation of water vapour is a process associated to heat release, the lapse rate significantly reduces. On average the atmospheric lapse rate is $\Gamma \simeq 6.5$ K km^{-1}. Sometime the lapse rate near the surface is negative; this phenomenon is known as thermal inversion and can result from several causes, like subsidence phenomenon when a large mass of warm air floats over a mass of colder air near the ground. From the top of troposphere up to 20 km the temperature is constant to a value of about 220 K (tropopause). In the stratosphere the temperature increases again up to 50 km, where the temperature value of about 273 K is reached. This temperature inversion is

connected with virtually absence of vertical mixing. At a height of about 50 km the high concentration of ozone layer is responsible of the strong UltraViolet (UV) radiation absorption, with the consequent stratospheric warming. At this height (stratopause) the temperature newly begins to decrease (mesosphere). The local minimum of temperature ($180 - 190$ K) define the mesopause. The thermosphere is the seat of a new temperature inversion, caused by the absorption of UV solar radiation, which originates the formation of ions (ionosphere). The last part of the atmosphere before the cosmic vacuum is the exosphere. In the upper part of the atmosphere the air is so thin that the few atoms and molecules are not enough to transfer any appreciable quantity of heat. However, because of the high kinetic energy of molecules, the exosphere is characterised by a kinetic temperature above 2000 K.

An example of profiles of atmospheric variables, given by a radiosounding measurement is given in Fig. 1.2. It is compared with "1976 U.S Standard Atmosphere", an idealised approximation of the average atmospheric conditions in function of geopotential height, used as reference. The equations to compute standard temperature and pressure as a function of geopotential height are given in Eqs. 1.18, 1.19.

$$\begin{cases} T = 288.15 - 6.5 \cdot h & h \leq 11 \text{ [km]} \\ T = 216.65 & 11 < h \leq 20 \text{ [km]} \\ T = 196.65 + h & 20 < h \leq 32 \text{ [km]} \\ T = 228.65 + 2.8 \cdot (h - 32) & 32 < h \leq 47 \text{ [km]} \\ T = 270.65 & 47 < h \leq 51 \text{ [km]} \end{cases} \tag{1.18}$$

$$\begin{cases} P = 101.325 \cdot (288.15/T)^{-5.255877} & h \leq 11 \text{ [km]} \\ P = 22.632 \cdot \exp\left[-0.1577 \cdot (h - 11)\right] & 11 < h \leq 20 \text{ [km]} \\ P = 5.4749 \cdot (216.65/T)^{34.16319} & 20 < h \leq 32 \text{ [km]} \\ P = 0.868 \cdot (228.65/T)^{12.2011} & 32 < h \leq 47 \text{ [km]} \\ P = 0.1109 \cdot \exp\left[-0.1262 \cdot (h - 47)\right] & 47 < h \leq 51 \text{ [km]} \end{cases} \tag{1.19}$$

No standard definition is available for the water vapour profile, because of its high variability in function of many parameter including height, horizontal and vertical dynamics, position latitude season and time of the day. The comparison of pressure show a very high similarity in the vertical profile values. The pressure in fact shows relevant changes only in local values, specially in surface ones, not in the vertical

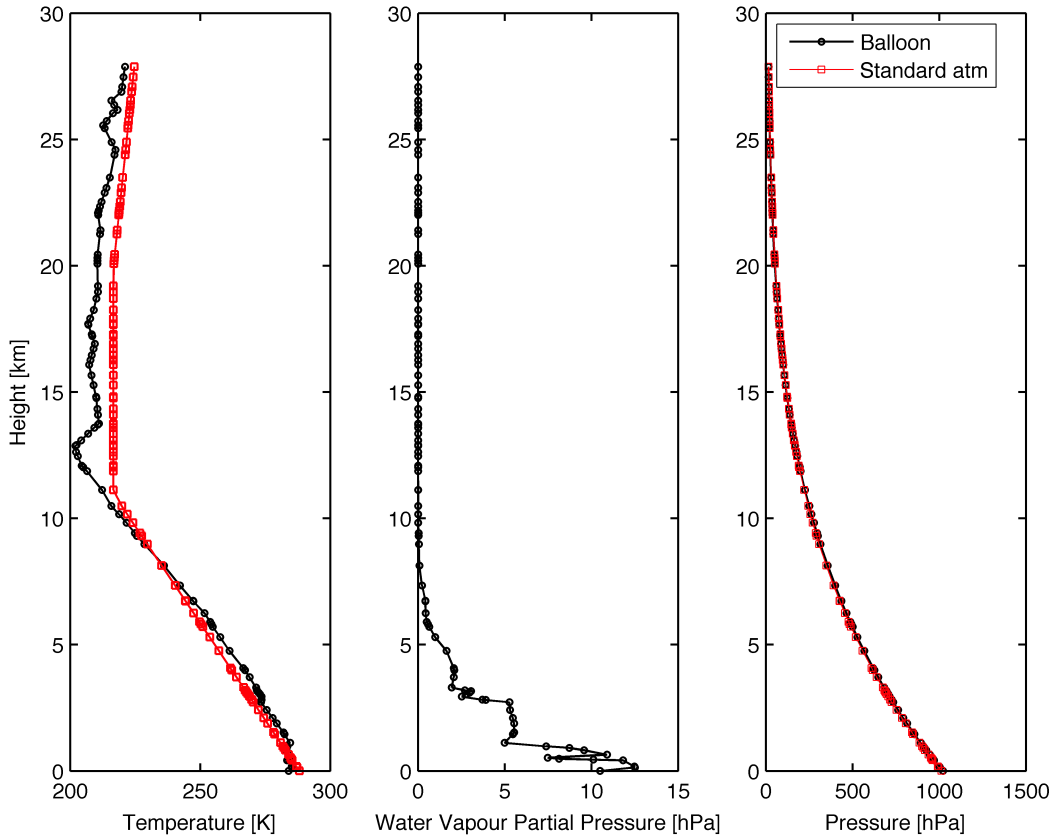


Figure 1.2: Balloon observations vs. standard atmosphere model.Balloon measurement are relative to Cagliari launch site on date 2011/01/07 at midnight.

pattern. The radiosonde temperature profile instead can differs a lot from the values of the standard atmosphere model. In the represented case a temperature inversion occurs near the surface level, this is quite evident in the showed profile, where the temperature increases in the first kilometer. The real height of troposphere is greater than 11 km with a minimum temperature value very lower than the standard one. Some fluctuations occurs in the tropopause, and in the stratosphere, but the values tend to converge to the model. Finally the observed water vapour pressure profile, shows a typical indented vertical variability, which prevent to profitably model this variable. As expected the concentration of water vapour is confined to the first kilometers of the troposphere.

1.2 Ionosphere

1.2.1 Classification of the ionospheric layers

The ionosphere is formed as a result of ionising (UV and X-ray) radiation coming from the sun; solar processes have effect in the photo-ionisation resulting in formation of ions positively charged (N_2^+, O_2^+, O^+, N^+, NO^+) and electrons. This ionisation strongly depends on solar activity and produces clouds of free electrons, which have a high day

variability, reaching maximum values in the first hours after local noon and minimum values in the early hours after midnight. Considering the seasonal variation the ionosphere is usually more stable in temperate zones, while can considerably fluctuate closer to the equator and the magnetic poles.

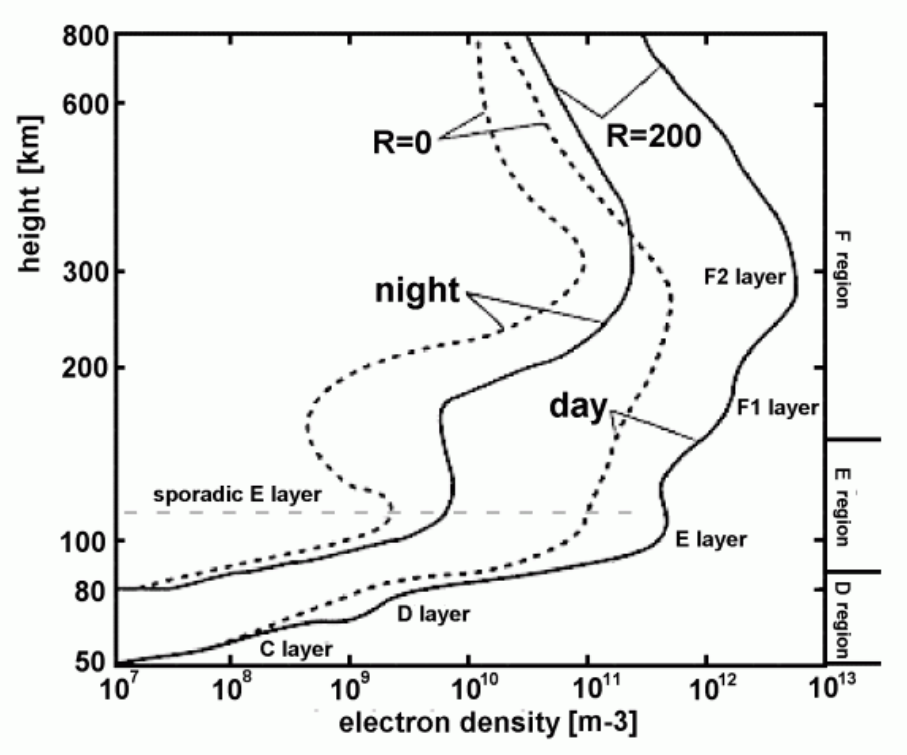


Figure 1.3: Typical electron density profiles of day and night in the ionosphere shown for low and high index (R) of solar activity ([6]).

Historically different regions have been defined for ionosphere depending on their chemical and physical composition, and their effects on radio signals at different frequencies (see Fig. 1.3):

- D $(50-90$ km): the lower part of the ionosphere. Normally this region is present only during daytime, when the sun activity and consequently the ionisation process are maximum. The D region is sometimes further divided in the C layer (with a maximum of ionisation at about 55 km) and D layer (with a maximum of ionisation at about 65 km).

- E $(90-150$ km): It is formed by a local maximum of electron density. Both the D region and the E region are characterised by a predominance of heavy ions (O_2^+ and NO^+). It is present also during nighttime.

- F_1 $(150-200$ km): mainly composed by light ions (NO^+). This layer has greater consistency during daytime and during summer, and almost vanishes during nighttime in winter.

- F_2 ($200 - 1000$ km) mainly composed by light ions (O^+ and H^+). It exhibits the highest concentration of electrons. After the absolute maximum the ions density decreases.

Above 1000 km a further ionised stratus (Protonosphere) is characterised by the presence of H^+ ions. The electron density of the protonosphere is almost constant from day to night, but it is noticeably affected by the presence of magnetic storms.

1.2.2 Model of ionospheric electron density

The models of the electron density in the ionosphere can be classified in two major groups: empirical and physical models. The first empirical models were developed to compute the parameter f_0F_2, the critical frequency[3] of the F_2 layer, which is proportional to the density of the peak electron density in the F_2 layer.

Physical models include as forcing functions the main physical processes involved in the creation of electrons and in the changes of electron density. Both empirical and physical models are based on large dataset of ionospheric observations, and consequently they have seasonal validity. The high day by day variability of the ionosphere is very difficult to be predicted and modeled.

The most diffuse model of the ionosphere is the Klobuchar model, which is taken as reference in the GNSS community also for its simplicity (only eight coefficients) so as to be directly broadcasted into the navigation message. These coefficients are sufficient to describe the global ionospheric status. Unfortunately they can be updated only once a day. The model assumes that the content of electrons is concentrated in a shell at 350 km of height. The intersection of the receiver satellite ray path with this shell is called Ionospheric Pierce Point (IPP) and it is used to account for the effects of ionosphere in GNSS. The Klobuchar model takes advantage of the fact that the average local daily behavior of the ionosphere can be easily expressed using only few terms. The daily trend is then approximated by a cosine function fit with a maximum at 14:00 local time (see § 1.4.3 for more details).

A different but more precise model of the Ionosphere has been developed (and it is constantly updated) by the International Reference Ionosphere (IRI) working group ([7]). The IRI model is an empirical model based on the available data sources for the ionospheric observation, as ionosondes, scatter radars, in situ instruments. For any given location, time and date, IRI provides monthly averages of the electron density, electron temperature, ion temperature, and ion composition in the altitude range from 50 km to 2000 km. As additional parameter, the Total Electron Content (TEC) is provided. The TEC is the integrated vertical content of free electron in the ionosphere

[3]The critical frequency is the highest frequency of a signal transmitted vertically that will reflect directly back to its transmission location.

(i.e number of electrons for unit surface):

$$TEC = \int_L N_e dl \tag{1.20}$$

where N_e is the electron density and L is the signal travel path. The last developed IRI model is the IRI-2007, described and freely available at [7]. An example of global TEC map retrieved from IRI-2007 and regional observation of f_0F_2 is given in Fig. 1.4.

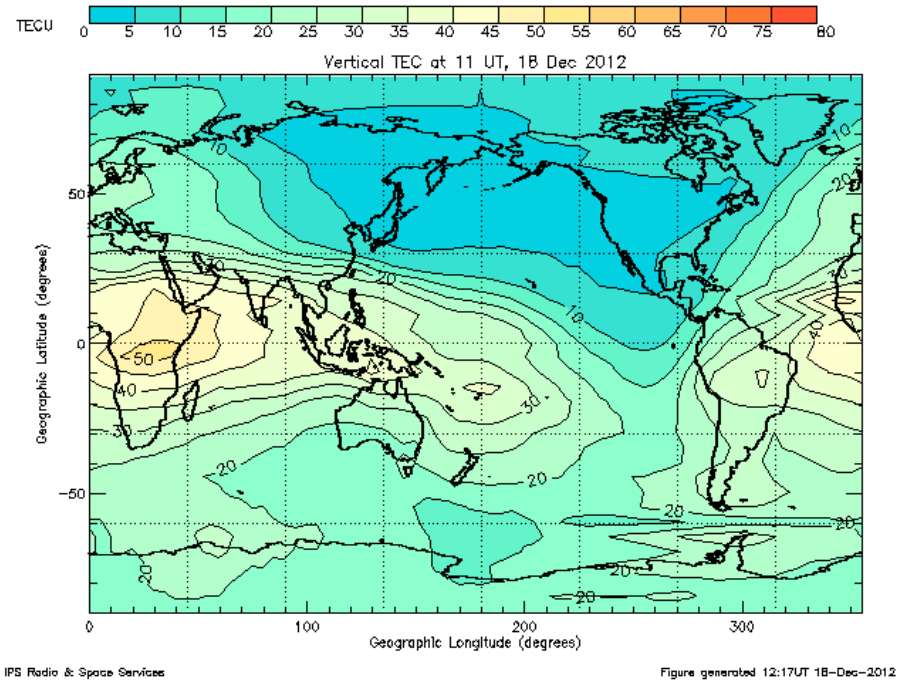


Figure 1.4: Global map of TEC ([8]).

The GNSS is very useful for the observation of the ionospheric TEC, as the navigation signal coming from the satellite is influenced by the distribution of electrical charges within the ionosphere. This interaction can be precisely modeled (§ 1.4.1) and used to reconstruct with good approximation also the 3D distribution of electron density in the ionosphere.

1.3 Effects of neutral atmosphere on radio signals

The main effects of troposphere in radio signal propagation are attenuation, scintillation, delay, bending. The present analysis is restricted on the effects that are directly connected with the operating techniques of GNSS system. As satellite positioning techniques are based on the travel time of the signal between the satellite and the receiver, and ranging techniques, only the refraction effect (signal delay and bending) will be considered and discussed in this work. Attenuation only concerns the amplitude of the signal, leaving intact the signal travel time; when the amplitude of a satellite signal at receiver level is below the MDS (Minimum Detectable Signal) of the receiver, the

satellite is simply not tracked from the receiver, thus reducing the number of observations at that epoch. Fortunately anywhere at the Earth surface there are more than 5 or 6 satellite simultaneously on view, with the result that even in the unlikely event of the loss of a satellite signal the system is anyway operating, and the receiver can compute its position. The scintillation only affects signals at very low elevation angles, with effects on amplitude and phase. The amplitude scintillation is a sudden amplitude variation with similar consequences as attenuation, in the positioning operations. The phase scintillation is a sudden (random) phase variation whose consequences in the worst case may imply the loss of signal tracking, which technically is referred as cycle slip, a jump in the GNSS carrier phase measurement (see [9] and [10] for more details).

Refraction is an effect on light or generally electromagnetic waves caused by transition from one medium to another, during propagation. It takes place with a double effect on electromagnetic waves: the first is a deviation of the trajectory from a straight line (bending) and the second is the speed change, namely a reduction, with respect to the speed in the free space (slowing). Refraction in the atmosphere is "complicated" by the continuous change of density of air, which decreases in function of height. The problem is commonly addressed by means of the quantity called refractive index n of a medium, which is linked to the speed of an electromagnetic medium propagating in it.

$$n = \frac{c}{v} \tag{1.21}$$

where c is the speed of light in vacuum, and v is the speed of light in the considered medium. The refractive index is introduced in the equation of propagation in space and time of the electric field:

$$E(\mathbf{r}, t) = E_0 e^{i[n(\mathbf{r})\mathbf{k}\cdot\mathbf{r} - \omega t]} \tag{1.22}$$

where E_0 is the electric field amplitude, $n(\mathbf{r})$ the refraction index in function of the position vector $\mathbf{r} = \mathbf{r}(x, y, z)$, $\mathbf{k}$ the wave vector ($|\mathbf{k}| = \frac{2\pi}{\lambda}$) , $\omega = 2\pi f$ the angular frequency.

1.3.1 Refraction in troposphere

As the quantity given by the difference between refractive indexes respectively in the atmosphere and in the vacuum $(n-1)$ is generally less than $3 \cdot 10^{-4}$, we use refractivity instead of refraction index to directly scale the quantities to the values that considerably affect the propagation mechanism. Refractivity is a quantity provided by the relationship:

$$N = (n - 1) \times 10^6 \tag{1.23}$$

In the atmosphere the concentration of gases like nitrogen, oxygen, argon, carbon dioxide is almost constant if compared with the water vapour concentration, whose

variation are much larger (both in space and time). In general water vapour molecules can give a very significant contribution on the refractivity of a signal, with a typical signature due to their dipole moment, thus the most commonly used refraction index formula is directly dependent on the concentration of water vapour.
The general formula used for the refractivity is:

$$N = K_1 \frac{P_d}{T} Z_d^{-1} + K_2 \frac{e_w}{T} Z_W^{-1} + K_3 \frac{e_w}{T^2} Z_W^{-1} \tag{1.24}$$

where:
K_1; K_2; K_3 are empirical constants (see Table. 1.2)
P_d [hPa] is the partial pressure of dry air
T [K] is the temperature
e_w [hPa] is the partial pressure of water vapour
Z_d^{-1} is the inverse of compressibility factor of dry air
Z_w^{-1} is the inverse of compressibility factor of water vapour
Compressibility factors (Z) are introduced to take into account non ideal behaviors of gases, considering real gas relationship: $PV = ZRT$, given pressure P, volume V, universal gas constant R and temperature T. Their derivation is beyond the scope of the present work, however the expressions for the compressibility are given by:

$$\begin{aligned} Z_d^{-1} &= 1 + P_d[57.97 \times 10^{-8}(1 + \frac{0.52}{T}) - 9.4611 \times 10^{-4} \frac{T_c}{T^2}] \\ Z_w^{-1} &= 1 + 1650 \frac{e_w}{T^3}[1 - 0.01317T_c + 1.75 \times 10^{-4} T_c^2 + 1.44 \times 10^{-6} T_c^3] \end{aligned} \tag{1.25}$$

T_c is the temperature in Celsius. Pressure and density of dry and wet air are related by the gas law:

$$\begin{aligned} P_d &= \rho_d \frac{R}{M_d} T Z_d \\ e_w &= \rho_w \frac{R}{M_w} T Z_w \end{aligned} \tag{1.26}$$

where:
$R = 8.31434$ kJ kmol^{-1} K^{-1} is the universal gas constant
$M_d = 28.9644$ kg kmol^{-1} is the molar weight of dry air
$M_w = 18.0152$ kg kmol^{-1} is the molar weight of water vapour
$\frac{M_w}{M_d} = 0.621977 \simeq 0.622$ as defined in 1.8
By substituting these equations in the refractivity formula (Eq. 1.24), introducing the total mass density $\rho_T = \rho_d + \rho_w$, the new coefficient $K_2' = K_2 - K_1 \frac{M_w}{M_d}$ and developing:

$$N = K_1 \rho_T \frac{R}{M_d} + K_2' \frac{e_w}{T} Z_W^{-1} + K_3 \frac{e_w}{T^2} Z_W^{-1} \tag{1.27}$$

In literature there are different sets of values for the refractivity coefficients (K_1,K_2,K_3), which have been experimentally obtained and compared over the years. In table 1.2 a list of the most diffuse values for these coefficients is given.

Autors	K_1 [K hPa^{-1}]	K_2 [K hPa^{-1}]	K_3 10^5 [K^2 hPa^{-1}]
Smith and Weinthraub (1953)	77.607 ± 0.013	71.6 ± 8.5	3.747 ± 0.031
Thayer et al. (1974)	77.604 ± 0.014	64.79 ± 0.08	3.776 ± 0.004
Hasegawa and Stokesberry (1975)	77.600 ± 0.032	69.40 ± 0.15	3.701 ± 0.003
Bevis et al. (1994)	77.60 ± 0.05	70.04 ± 2.2	3.739 ± 0.012
Rüeger (2002)	77.6890 ± 0.0094	71.2952 ± 1.3	3.75463 ± 0.0076

Table 1.2: Survey of refraction index coefficients.

A simplified relationship used for refractivity in the computation of electromagnetic wave propagation in troposphere can be obtained neglecting the second term contribution and tuning the coefficients accordingly. An example is given by:

$$N = \frac{77.6}{T} \times \left(P + 4810\frac{e_w}{T}\right) = 77.6\frac{P}{T} + 3.732 \times 10^5 \frac{e_w}{T^2} \tag{1.28}$$

Eq. 1.28 is not usable for our purposes of estimation of atmospheric parameters, due to the inaccuracy of the estimation model whose relative error can reach 0.5%.

The refractivity can be seen as composed by two terms: the hydrostatic refractivity N_h and the wet refractivity N_w:

$$\begin{aligned} N_h &= K_1 \rho_T \frac{R}{M_d} = 222.77 \rho_T \\ N_w &= K_2' \frac{e_w}{T} Z_W^{-1} + K_3 \frac{e_w}{T^2} Z_W^{-1} \end{aligned} \tag{1.29}$$

Correspondingly also the signal delay, as it is the integration of the total refractivity, can be decomposed in two terms, called hydrostatic and wet delays respectively.

The origin of the name hydrostatic is due to the approximation (hydrostatic atmosphere) used during the processing of this component. Some details about the processing algorithms used to retrieve the water vapour content starting from this two delay components are given in next sections.

1.3.2 Hydrostatic component

The hydrostatic refractivity varies in function of atmosphere density following the formula of Eq. 1.29. The Zenith Hydrostatic Delay (ZHD)[4] can be expressed as:

$$ZHD = \int_0^{TOT} N_h(z) \times 10^{-6}\, dz = 222.77 \times 10^{-6} \int_0^{TOT} \rho_T(z)\, dz \tag{1.30}$$

where Top Of Troposphere (TOT) is the upper limit of the portion of atmosphere where this model of refractivity can be used. Theoretically the upper integration limit should be the distance between the satellite and the Earth surface, but beyond the troposphere refractivity as expressed by Eq. 1.24 is negligible and take over the ionospheric effects (see § 1.4).

[4]Zenith stands for along the vertical path.

Using the hydrostatic equation (Eq. 1.3) the surface pressure P_0 can be expressed as the integral of the product of the density ρ and the acceleration of gravity g from the surface level ($z = 0$) to the top of the troposphere ($z = TOT$):

$$P_0 = \int_0^{TOT} \rho_T(z) g(z)\ dz = g_{eff} \int_0^{TOT} \rho_T(z)\ dz \tag{1.31}$$

where:

$$g_{eff} = \frac{\int_0^{TOT} \rho_T(z) g(z)\ dz}{\int_0^{TOT} \rho_T(z)\ dz} \tag{1.32}$$

is the effective gravitational constant.
Defining:

$$G = \frac{g_{eff}}{9.784} = (1 - 0.002626 \cos(2lat) - 0.00028 z_0) \tag{1.33}$$

where lat is the receiver latitude, z_0 is the receiver height.
The hydrostatic delay can be expressed as:

$$\begin{aligned} ZHD &= 222.77 \times 10^{-6} \frac{P_0}{9.784 G} = \\ &= 222.77 \times 10^{-6} \frac{P_0}{9.784(1 - 0.002626 \cos(2lat) - 0.00028 z_0)} \end{aligned} \tag{1.34}$$

with P_0 expressed in Pascal.

1.3.3 Wet component

Similarly the Zenith Wet Delay (ZWD) can be derived from the wet refractivity formula (Eq. 1.29):

$$\begin{aligned} ZWD &= 10^{-6} \times \int_0^{TOT} N_w(z)\ dz = 10^{-6} K_2' Z_W^{-1} \int_0^{TOT} \frac{e_w}{T}\ dz + \\ &+ 10^{-6} K_3 Z_W^{-1} \int_0^{TOT} \frac{e_w}{T^2}\ dz = \\ &= 10^{-6} Z_W^{-1} \int_0^{TOT} \frac{e_w}{T^2}\ dz \left[K_3 + K_2' \frac{\int_0^{TOT} \frac{e_w}{T}\ dz}{\int_0^{TOT} \frac{e_w}{T^2}\ dz} \right] \end{aligned} \tag{1.35}$$

Following [11] the weighted mean temperature can be obtained from the formula:

$$T_m = \frac{\int_0^{TOT} \frac{e_w}{T}\ dz}{\int_0^{TOT} \frac{e_w}{T^2}\ dz} \tag{1.36}$$

So the zenith wet delay can be rewritten as:

$$ZWD = 10^{-6} Z_W^{-1} \int_0^{TOT} \frac{e_w}{T^2}\ dz \left[K_3 + K_2' T_m \right] =$$

and using Eq. 1.36

$$\begin{aligned} &= 10^{-6} Z_W^{-1} \left[K_3 + K_2' T_m\right] \frac{\int_0^{TOT} \frac{e_w}{T}\, dz}{T_m} = \\ &= 10^{-6} Z_W^{-1} \left[K_2' + \frac{K_3}{T_m}\right] \int_0^{TOT} \frac{e_w}{T}\, dz = \\ &= 10^{-6} Z_W^{-1} \left[K_2' + \frac{K_3}{T_m}\right] \frac{R}{M_w} Z_W \int_0^{TOT} \rho_w\, dz = \\ &= 10^{-6} \left[K_2' + \frac{K_3}{T_m}\right] \frac{R}{M_w} IWV \end{aligned} \tag{1.37}$$

The Integrated Water Vapour (IWV) gives the total amount of water vapour that a signal from the zenith direction would encounter. Precipitable Water (PW) is the IWV scaled by the density of water and is referred to as the total atmospheric water vapour contained in a vertical column of unit cross-sectional area, commonly expressed in terms of the height to which that water substance would stand if completely condensed and collected in a vessel of the same unit cross section. PW and IWV are generally expressed as mm and kg m^{-2} respectively.

$$IWV = \int_0^{TOT} \rho_w\, dz = \rho PW \tag{1.38}$$

where ρ is the density of water. The direct formula to retrieve PW from the ZWD is:

$$PW = 10^6 \times \frac{1}{\rho} \frac{M_w}{R} \frac{1}{\left(K_2' + \frac{K_3}{T_m}\right)} ZWD \tag{1.39}$$

1.4 Effect of ionosphere on radio signals

The ionosphere interacts with electromagnetic waves differently depending on the frequency. In the High Frequencies (HF) range (3 to 30 MHz) the ionosphere allows to establish long range communications, mainly through the mechanisms of signal reflection. In fact all the signal at frequency below the so called plasma or critical frequency (Eq. 1.50) of the ionosphere are reflected. Typical values of the ionosphere plasma frequencies are on the range of 0.1 to 30 MHz. At lower frequencies the D region reflects most of the electromagnetic radiation. By increasing the frequency the lower layer become transparent and the electromagnetic waves are reflected by the higher shells, typically the E and F regions at HF frequencies.

Because of the number of free electrons, the ionosphere and the overlying plasmasphere (or protonosphere) behave as a dispersive medium that significantly affect the propagation of high frequency signals. This also occurs at GPS frequencies ($f_{L1} = 1575.42$ MHz, $f_{L2} = 1227.60$ MHz), where the most significant contribution is generated within the regions F_2 and the plasmasphere. The velocity of propagation varies as a function of frequency; the modulating signal is delayed in proportion to

the number of free electrons. The entity of that delay is (at first order) inversely proportional to the square of the carrier frequency. Also the phase of the carrier has a variation of the same amount but of opposite sign, which occurs as a phase advance.

Users interested in positioning must correct the pseudorange measurements taking into account for such delays. The most simple correction can be applied also for single-frequency receivers operating at f_{L1}, using a daily model of the ionosphere delay (Klobuchar model [12]), whose parameters are broadcasted within the GPS signal (see § 1.4.4 for more details). A second technique for dual frequency receivers measure the signal at both frequencies, and, through a combination of measures of the times of arrival of f_{L1} and f_{L2}, a direct algebraic solution can be applied (§ 2.3.4)). A third technique uses a real-time update. It is an example of the system Wide Area Differential GPS (WADGPS) that uses a network of local permanent stations conveniently located and calculates the ionospheric delay (valid only locally) using differential techniques.

Eq. 1.22 is the general equation of a propagating electromagnetic wave and it is a general solution of wave equation:

$$\nabla^2 E - \left(\frac{1}{v_p}\right)^2 \frac{\partial^2 E}{\partial t^2} = 0 \tag{1.40}$$

where v_p is is the speed of cycle wave, the phase velocity, defined as the ratio between the angular frequency (ω) and the wave number (k):

$$v_p = \frac{\omega}{k} \tag{1.41}$$

The phase velocity is not necessarily the speed at which energy or information propagates, which instead is correctly given by the group velocity:

$$v_g = \frac{\partial \omega}{\partial k} \tag{1.42}$$

By introducing the refraction index of the medium in which the wave propagates, as a function of relative electric permittivity (ϵ_r) and relative magnetic permeability (μ_r):

$$n = \sqrt{\epsilon_r \mu_r} \tag{1.43}$$

the propagation velocity of the phase of the wave can be expressed as a function of the propagation velocity in the free space (c):

$$v_p = \frac{c}{n} = \frac{1}{\sqrt{\epsilon \mu}} \tag{1.44}$$

By comparison of Eqs. 1.41 and 1.44 it comes out that: $\omega = \frac{kc}{n}$; and the derivative with respect to k is:

$$\frac{\partial \omega}{\partial k} = \frac{c}{n} - \frac{kc}{n^2}\frac{\partial n}{\partial k} \tag{1.45}$$

The first term of Eq. 1.45 is coincident with the definition of group velocity, so:

$$v_g = v_p \left(1 - \frac{k}{n}\frac{\partial n}{\partial k}\right) \tag{1.46}$$

Eq. 1.46 can be modified by considering the expressions: $\partial n = -\frac{c}{v_p^2}\partial v_p$ and $\partial k = -\frac{2\pi}{\lambda^2}\partial\lambda$. The result is the expression:

$$v_g = v_p - \lambda\frac{\partial v_p}{\partial\lambda} \tag{1.47}$$

In addition considering $\frac{\partial\lambda}{\lambda} = -\frac{\partial f}{f}$, Eq 1.47 can be rewritten as:

$$v_g = v_p + f\frac{\partial v_p}{\partial f} \tag{1.48}$$

If the electromagnetic wave is propagating in the vacuum, the group and the phase velocity are identical and are equal to the speed of light in the free-space. In such a case the medium is called non-dispersive. If group and phase velocities are different, the medium is said dispersive, as the ionosphere is, and two different indexes are introduced for phase and group respectively, n_p and n_g: $v_p n_p = c$; $v_g n_g = c$.
Differentiating v_p with respect to f: $\partial v_p = -\frac{c}{n_p^2}\partial n_p$

$$n_g = \frac{n_p^2}{n_p - f\frac{\partial n_p}{\partial f}} = n_p^2\frac{1}{n_p}\frac{1}{1 - \frac{f}{n_p}\frac{\partial n_p}{\partial f}}$$

Using the Taylor expansion $(1-x)^{-1} = 1 + x - x^2$ limiting at first order:

$$n_g = n_p + f\frac{\partial n_p}{\partial f} \tag{1.49}$$

The dispersive effects of the ionosphere give as effect differences between phase and group velocities, and therefore different effects on the *pseudorange* and *phase* observables (§ 1.4.3).

1.4.1 Refraction in Ionosphere

The most commonly used relationship for the refraction index in the ionosphere is the Appleton-Hartree formula, with a validity restricted to a number of hypothesis.
Some hypothesis concern the properties of the signal which is assumed to be a plane wave with small amplitude and simple harmonic; some other are on the properties of the medium that is assumed electrically neutral, with charge distributed statically and uniformly (density N_e) resulting in absence of space charges and influenced by an uniform external magnetic field. In addition the electron collision probability is assumed independent on the electron energy, the thermal motions of electrons is assumed negligible (cold plasma).

$$n^2 = (\mu - i\chi)^2 = 1 - \frac{X}{1 - iZ - \frac{Y_T^2}{2(1-X-iZ)} \pm \left[\frac{Y_T^4}{4(1-X-iZ)^2} + Y_L^2\right]^{\frac{1}{2}}} \tag{1.50}$$

where:

μ and χ are the real and imaginary component of the refraction index respectively,

$$X = \frac{N_e e^2}{4\pi^2 \epsilon_0 m f} = \left(\frac{f_p}{f}\right)^2$$

$$Y_T = \frac{e\mu_0 H_T}{2\pi m f} = \frac{f_H \sin\theta}{f}$$

$$Y_L = \frac{e\mu_0 H_L}{2\pi m f} = \frac{f_H \cos\theta}{f}$$

$$Z = \frac{\nu}{2\pi f}$$

being:

ϵ_0 = electric permittivity of free space = $8,854 \times 10^{-12}$ F m^{-1}

μ_0 = magnetic permeability of free space = $4\pi \times 10^{-7}$ H m^{-1}

e = electron charge = $-1,602 \times 10^{-19}$ Coulomb

m = electron mass = 9.107×10^{-31} kg

f = signal frequency [Hz]

N_e = electron density [electrons m^{-3}]

$f_p = \sqrt{\frac{N_e e^2}{4\pi^2 \epsilon_0 m}}$ plasma frequency [Hz]

H = geomagnetic field amplitude [A m^{-1}], subscription T and L are referred to transversal and longitudinal component of the field respect to the propagation direction

ν = electron collision frequency varying in function of the electron distribution (for example at a level of 100 km $\nu = 1,75 \times 10^3$ Hz)

f_H = gyrofrequency, the measurement of interaction between electrons moving in a circular motion in the ionosphere and the magnetic field with energy dispersion; more the frequency of the signal is close to the gyrofrequency of the electron, more is the energy absorption. f_H is of the order of 1.4MHz.

θ = angle between magnetic field vector and the wave direction

$\pm$ provide different solutions for the complex refractive index, depending on wave modes[5]

[5]In an unmagnetised plasma an electromagnetic wave behaves simply as a light wave modified by the plasma medium. In a magnetised plasma on the contrary the situation is different and we can have two wave modes perpendicular to the field, the O and X modes, and two wave modes parallel to the field, the R and L ones. For propagation perpendicular to the magnetic field ($k \perp H_0$), the '+' sign is due to the "ordinary" (O) mode and the '-' sign due to the "extraordinary" (X) one. For propagation parallel to the magnetic field ($k \parallel H_0$), the '+' sign is due to a left-hand circularly polarised mode (L), and the '–' sign due to a right-hand circularly polarised mode (R).

In Eq 1.50 the square of the refraction index and consequently the real part of the refraction index is less than one. Therefore the phase speed is greater than the light speed on the free space. This does not violate the principles of special relativity, as the speed of signal is the group velocity which, as seen in the previous paragraph, will be different from the phase velocity, and it will remain always lower than the speed of light in the vacuum.

1.4.2 Refraction index approximations

Some approximation can be made to simplify Eq. 1.50. If the electron collision frequency is negligible with respect to the signal frequency ($f \gg \nu$), the term Z can be neglected ($Z \simeq 0$), we have:

$$n^2 = 1 - \frac{X(1-X)}{1 - X - \frac{1}{2}Y^2 \sin^2\theta \pm \left[\frac{1}{4}Y^4 \sin^4\theta + Y^2 \cos^2\theta (1-X)^2\right]^{\frac{1}{2}}} \tag{1.51}$$

where $Y = \frac{f_H}{f}$. Considering that the X e Y at GNSS working frequencies are very small ($X < 4,4 \times 10^{-5}$) and ($Y_T, Y_L < 1 \times 10^{-3}$), the denominator of the fraction can be developed using Taylor expansion:

$$\begin{aligned} n^2 &= 1 - X(1-X) \times \\ &\times \left\{1 + X + \frac{1}{2}Y^2 \sin^2\theta \mp \left[\frac{1}{4}Y^4 \sin^4\theta + Y^2 \cos^2\theta (1-X)^2\right]^{\frac{1}{2}} + ...\right\} \end{aligned} \tag{1.52}$$

Neglecting terms less than 10^{-9}, and developing the term within the square brackets, we obtain:

$$n^2 = 1 - X(1-X)\left[1 + X \mp Y\cos\theta\right] \tag{1.53}$$

Developing and newly neglecting terms less than10^{-9}, we can write:

$$n = \sqrt{1 - X \pm XY\cos\theta} \tag{1.54}$$

Using the Taylor expansion $\sqrt{1-A} = 1 - \frac{A}{2} + \frac{A^2}{8} + o(A^2)$ and neglecting again terms less than 10^{-9}, we finally have:

$$n = 1 - \frac{X}{2} \pm XY\cos\theta - \frac{X^2}{8} \tag{1.55}$$

Eq 1.53 gives the same result of Eq 1.50 making the hypothesis of propagation longitudinal with respect to terrestrial magnetic field ($\sin\theta \ll 1$), but in this case no assumption has been made on the direction of the magnetic field vector and the signal propagation.

Eq 1.55 is dependent on the terrestrial magnetic field (term Y) and needs a model of magnetic field to be used. Limiting the expansion to first order, the model accuracy decreases, but we can verify that it is in any case better than 1 %:

$$n = 1 - \frac{X}{2} = 1 - \frac{1}{2}\frac{N_e e^2}{4\pi^2 \epsilon_0 m}\frac{1}{f^2} \tag{1.56}$$

The first order approximation is the most used because a standard dual frequency receiver allows to correct ionosphere delay (§ 2.3.4). However there are some other more precise approximations of the ionosphere refractive index, as an example here a formula of the refraction expressed as a power series (see [4] for more details):

$$n = 1 + \frac{c_2}{f^2} + \frac{c_3}{f^3} + \frac{c_4}{f^4} + ... \tag{1.57}$$

1.4.3 Group delay and phase advance

Eq 1.56 can be rewritten as:

$$n_p = 1 - \frac{A}{f^2} \tag{1.58}$$

where: $A = \frac{N_e e^2}{8\pi^2 \epsilon_0 m}$. Using Eq 1.49 we obtain:

$$n_g = 1 + \frac{A}{f^2} \tag{1.59}$$

Eqs. 1.58 and 1.59 give the approximation of the signal phase and group refractions in the ionosphere respectively. Expressing in terms of time delay (group delay and phase advance), we have:

$$Group\ Delay = \frac{1}{c}\int_L (n_g - 1)\partial l = \frac{1}{c}\int_L \frac{A}{f^2}\partial l \tag{1.60}$$

$$Phase\ advance = \frac{1}{c}\int_L (1 - n_p)\partial l = \frac{1}{c}\int_L \frac{A}{f^2}\partial l \tag{1.61}$$

This two terms have the same value. They are usually denoted by $\Delta T_{Ion}(f)$. The A term has constant value varying the frequency, and depends on the free electron content, on the ionospheric portion travelled by the GNSS signal.

$$\Delta T_{Ion}(f) = \frac{B}{f^2} \tag{1.62}$$

where B is the integral of the term A along the signal path divided by the velocity of light in the vacuum, and therefore it is constant varying the frequency.

Introducing the expression of Eq. 1.20 in Eq 1.62 the constant B can be expressed in function of TEC:

$$\Delta T_{Ion}(f) \simeq C \times \frac{TEC}{f^2} \tag{1.63}$$

where the constant value of $C = \frac{e^2}{8c\pi^2\epsilon_0 m} \simeq 40.28$. Considering higher order approximations (e.g Eq. 1.57) the terms become:

$$\Delta T_{Ion}(f) \simeq C \times \frac{TEC}{f^2} + \frac{S}{f^2} + \frac{R}{f^3} \tag{1.64}$$

where $S = \frac{7527}{2} c \int_L N_e H \| \cos\theta \| dl$ and $R = \frac{2437}{3}\int_L N_e^2 dl + \frac{4.74}{3} 10^{22} \int_L N_e H^2 (1 + \cos^2\theta) dl$.

1.4.4 Ionospheric delay computations in single frequency receiver

The value expressed in Eq. 1.62 can be computed with very high precision using both GNSS frequency and taking advantage of the dispersive properties of the ionosphere (§ 2.3.4). If a dual frequency receiver is not available, the ionosphere delay at the f_{L1} frequency can be estimated using the broadcasted parameters of the Klobuchar model. The slant ionosphere delay is computed in function of the vertical delay, due to the electron density at the IPP, by using an obliquity factor. The daily variability is modeled by using a cosine fit function:

$$\begin{aligned} \Delta T_{iono} &= F \times \left\{ DC + A\cos\left[2\pi\frac{(t-\phi)}{P}\right]\right\} \quad (day) \\ \Delta T_{iono} &= F \times \{DC\} \quad (night) \end{aligned} \tag{1.65}$$

where F is the slant factor used to correct with respect to the zenith value by using the elevation θ_e of the satellite receiver direction ($F = 1.0 + 16.0 \times (0.53 - \theta_e)^3$), DC is the offset term ($DC = 5 \times 10^{-9} sec$); ϕ is the phase of the peak (14:00 local time 50400 sec). The values of A and P are:

$$\begin{aligned} A &= \sum_{n=0}^{4} \alpha_n lon_m^n \\ P &= \sum_{n=0}^{4} \beta_n lon_m^n \end{aligned} \tag{1.66}$$

α_n and β_n are the coefficients broadcasted in the GPS navigation message and lon_m is the geomagnetic latitude, computed as follows, starting from the geodetical latitude (lon_u) and longitude (lat_u), the elevation (θ_e) expressed as semicircles[6] and the azimuth value (θ_a). As first step the Earth-centred angle (semicircles) is computed:

$$\psi = \frac{0.0137}{\theta_e + 0.11} - 0.022 \tag{1.67}$$

Then the latitude (lon_i) of the IPP is :

$$lon_i = lon_u + \psi \cos\theta_a \tag{1.68}$$

If $lon_i > +0.416$ then $lon_i = +0.416$. If $lon_i < -0.416$ then $lon_i = -0.416$ The longitude of the IPP is:

$$lat_i = lat_u + \frac{\psi \sin\theta_a}{\cos lon_i} \tag{1.69}$$

Finally the geomagnetic latitude is:

$$lon_m = lon_i + 0.064\cos(lat_i - 1.617) \tag{1.70}$$

The local time at the IPP is:

$$t = 4.32 \times 10^4 lat_i + GPS_{time} \tag{1.71}$$

If $t > 86400$ then $t = t - 86400$. If $t < 0$ then $t = t + 86400$.

[6]The semicircle measurements unit is used because the broadcasted coefficients contain the semicircle unit.

1.5 Bending effect in the atmosphere

The refraction of an electromagnetic signal occurs with two main effects: the slowing effect discussed in the previous chapters; and the change of propagation direction described in this chapter. The two terms can be analytically divided starting from a general definition of the delay (δt) of signal traveling through a given medium, with respect to the straight propagation in the vacuum:

$$\begin{aligned} \delta t &= \int_L \frac{1}{v(l)} dl - \int_G \frac{1}{c} dl = \\ &= \int_G \frac{n(l)}{c} dl + \int_{L-G} \frac{n(l)}{c} dl - \int_G \frac{1}{c} dl = \\ &= \frac{1}{c} \int_G [n(l) - 1] dl + \int_{L-G} \frac{n(l)}{c} dl = \end{aligned} \tag{1.72}$$

where L is the optical path of the signal, G is the geometrical distance, v and c the signal speed in the medium and in the vacuum respectively. In Eq. 1.72 the first term is the effect introduced by the slowing, the second term is the bending effect. This two effects are present both in troposphere and in ionosphere. As the ionosphere signal delay (Eq. 1.62) can be directly computed by using a combination of the two GNSS signal frequencies, and the resulting value includes both terms only the troposphere bending must be taken into account. The only contribution that would remain to consider would be the difference of the optical paths at the two frequencies, used when the ionosphere delay is computed (§ 2.3.4), but this is negligible, especially if we consider elevation angles greater than $15 - 20\,\text{deg}$ (see for instance [13]).

Assuming a vertically stratified atmosphere with a radial symmetry, the bending affecting a radiowave signal can be computed by applying the Snell's Law for each layer. Starting from the classical formulation of the Snell's Law valid for a planar geometry, with reference to Fig. 1.5:

$$n_i \sin \alpha_i = n_{i+1} \sin \alpha_{i+1} = n_i \sin \psi_i \tag{1.73}$$

By considering a spherically layered atmosphere geometry and applying the sine law to the triangle MP_iP_{i+1}[7] the following formula is obtained:

$$\frac{r_i}{\sin \psi_i} = \frac{r_{i+1}}{\sin(\pi - \alpha_i)} = \frac{r_{i+1}}{\sin \alpha_i} \tag{1.74}$$

By expressing $\sin \psi_i = \frac{r_{i+1}}{r_i \sin \alpha_i}$ from Eq. 1.74 and substituting in Eq. 1.73 the Snell law in polar coordinates (also known as Bouguer's formula)

$$n(r) r \sin \alpha(r) = constant = a \tag{1.75}$$

where r is the distance from the centre of mass M of the Earth, $\alpha(r)$ is the angle between the direction of r and the tangent to the raypath, $n(r)$ is the index of refraction at r.

[7] The arc P_iP_{i+1} in the layer n_i is aprroximated as a straight path.

Both n and α change only in function of the ray r according with the spherical symmetry hypothesis. The a constant can be determined by considering $r_0 = R_{max}$ over the top of the troposphere, where the refraction index n is reasonably approximated by 1, and the α angle is given by the geometry of the system satellite-receiver.

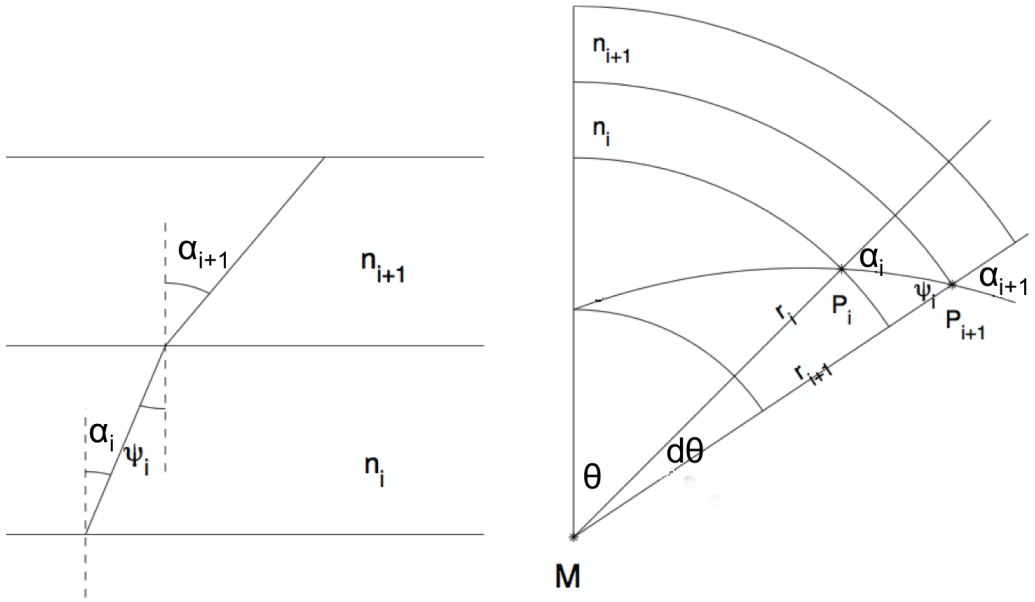


Figure 1.5: Geometry for the Snell's Law: planar (left panel) and spherical (right panel) ([14]).

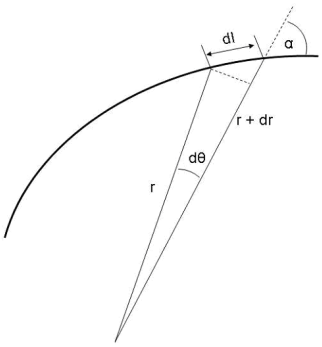


Figure 1.6: Infinitesimal optical path.

In order to compute the bending effect, the difference between the optical path of the signal in the atmosphere and the geometric distance (i.e the optical path for a signal traveling in the vacuum) must be evaluated. From Fig. 1.6 it can be easily deduced that $dl = \sqrt{dr^2 + r^2 d\theta^2}$ and $dr = dl \cos\alpha$. Combining it with the trigonometric relationship: $\cos\alpha = \sqrt{1 - \sin^2\alpha}$, and considering Eq. 1.75, it can be obtained:

$$dl = \frac{rn(r)}{\sqrt{r^2 n^2(r) - a^2}} dr \tag{1.76}$$

The signal optical path is given by the integral of Eq 1.76 between $r = r_e$ (Earth radius at the receiver) and the $r = r_{TOT}$ (height of the troposphere).

Chapter 2

The GNSS

2.1 The GPS satellite navigation system

Satellite navigation represents the last done step in the evolution of navigation, started from celestial observation used by ancient sailors and come down to the use of radio signals transmission and detection. Satellite navigation was born in the early 1960s by an initiative of U.S. Navy with the aim of creating a system for precise navigational purposes. The system was called Navy Navigation Satellite System (NNSS) or, more commonly, TRANSIT and was developed by Johns Hopkins Applied Physics Laboratory (APL) that starting from data coming from the first artificial satellite Sputnik I (Russian satellite) showed a method to track the satellite orbit using the doppler signal coming from the satellite and detected from known points on the Earth surface. The inversion of the problem led to the first satellite radio navigation system able to determine the position of the user by knowing the doppler shift of satellite signals and orbital parameters. The TRANSIT navigation system was developed for military purposes, mainly for submarine navigation. After the experience of TRANSIT, during the TIMATION (TIMe/navigATION) program of Naval Research Laboratory's (NRL's) Naval Center for Space Technology (NCST) the satellites where equipped with accurate oscillators (quartz oscillator in the initial project phase and atomic clocks in the final one) controlled and synchronised by ground master stations. The program, which started in 1967 and proved that a system using a passive ranging technique, combined with highly accurate clocks, could provide the basis for a new and revolutionary navigation system with three-dimensional coverage (longitude, latitude, and altitude) throughout the world. The results of the TIMATION program gave the basis for the development of the first satellite-based passive radio navigation system, the Navigation System with Time and Ranging (NAVSTAR), known as GPS. The GPS functioning principle in a nutshell is the following: when all satellite transmitters and the receiver are time synchronised, knowing the position and the signal transmission time (contained in the satellite signal) relative to three or more satellites, it is possible to analytically set and solve a system of three equations with the receiver coordinates as unknowns. The equa-

tions system is the result of a linearization process, as shown in § 2.3. For cost saving, the GPS receivers are not equipped with atomic clocks, so they aren't synchronised with the satellites; the time is thus introduced as further unknown and at least four satellite are needed to set and solve a system of four equations.

The GPS design started in 1973 and the system became fully operational in 1993. The main services provided by the GPS system are user position and speed vector determination and the time synchronisation of the global scale. The GPS system has been designed and developed from U.S. military departments for military and civilian users but the performance of the system can be degraded for reason of military safety, using the so called "Selective Availability" (SA). Since May 1^{th} 2000 the SA has been turned off, and the full precision of GPS system is available for all GPS users.

The architecture of the system is composed by space segment, control segment and users segment.

The space segment is the constellation of satellites, each sending a signal modulated with a code (unique and different for each satellite) combining information on time synchronisation and space vehicle position determination.

The Operational Control Segment (OCS) performs the tracking of space vehicles computing, monitoring and adjusting their position, the monitoring of space vehicle clock offsets and drifts in order to maintain synchronisation among the satellite clocks, and finally the updating of the navigation signal sent by the satellites. The control segment is composed by the Master Control Station (MCS) and five Monitor Stations (MS), some of them equipped with Ground Antennas (GA).

The User Segment consists of the final users of the GPS system. The hardware of user receivers detect the GPS satellite signals and give the position by means of the installed software.

The GPS system is designed to contemporarily provide two different services: Standard Position Service (SPS) and Precise Positioning Service (PPS), by equipping the system with two different type of signal sent by the satellites, called coarse (or clear) acquisition (C/A) code and precision (P) code respectively.

The system is designed to ensure the correct positioning in every place on the Earth. Positioning, as explained above, is possible where at least four satellites of the constellation are visible from the user position. The GPS system allows the user to receive simultaneously data from more than four satellite (up to ten in some cases). This redundancy is essential when the computation is not limited to the four unknowns which constitute the space-time position, but involves the other uncertainties: the bias and offset given by noise sources.

2.2 GPS constellation

2.2.1 Space segment structure

The Space Segment consists of the GPS satellite constellation. The minimum number ensuring the proper system working is 24, but a number of backup satellites increase the efficiency of the system. At the moment (2013) there are 31 operational satellites. The satellites move on 6 orbital planes, each of them containing at least 4 slots where satellites can be arranged equidistantly. The orbital planes have an inclination of 55 degrees with respect to the equatorial plane and are rotated in the equatorial plane by 60 degrees against each other. This geometry ensures total global coverage so that at least four satellites are simultaneously visible, any time and from any place, allowing to take advantage of the positioning service. The GPS orbits are Medium Earth Orbit (MEO), near circular, with a radius of about 26560 km, resulting in an eight of 20200 km above the mean Earth surface. The revolution time is half a sidereal day, precisely corresponding to 11 hours and 58 minutes. The speed of GPS satellites is about 3.9 km s^{-1}.

2.2.2 Brief history of GPS constellation

The constellation of first generation (block I) GPS satellites, consisting in 10 satellites was launched from 1978 to 1985. These satellites were experimental and are unused since many years. This series of satellites had an orbit inclination of 63 degrees against the equator. They constituted the GPS Demonstration system and reflected various stages of system development. The satellite weight was 850 kg and one cesium and two rubidium atomic clocks were on board of the space vehicles.

The second generation (block II), made of 9 satellites constellation was launched from 1989 to 1990 to provide 14 days of operation without contact from the Control Segment (CS). The following generation of satellites (block IIA), consisting of 19 satellites, was launched between 1990 and 1997. These satellites were designed to provide 180 days of operation without contact from the control segment. During the 180 day autonomy, degraded accuracy was evident in the navigation message. The block II and IIA satellites are equipped with two rubidium and two cesium atomic clocks and the weigh is more than 1500 kg. They have the Selective Availability (SA) and Anti-Spoof (A-S) capabilities.

When the GPS system became operational in 1993 the constellation of 24 satellites was composed of block I/II/IIA. Then, 12 satellites of block IIR where successfully launched between 1997 and 2004. They were called "Replenishment" satellites. Each satellite weights 2030 kg at launch and 1080 kg once on orbit. An important innovation was the capability of autonomously navigate (AUTONAV) themselves, by creating the

50 Hz navigation signal. This includes the ability to determine its own position by performing inter-satellite ranging with other IIR space vehicles.

Further 8 satellites were launched between September 2005 and August 2009, making the IIR-M (Modernisation) block. Their capabilities include developmental military-use-only M-code on the L1 and L2 signals and a civil code on the L2 signal, known as the L2C signal.

Both block IIR and block IIR-M are equipped with three rubidium atomic clocks. Their extreme precision with a drift of ± 1 second in 1 million years, is absolutely necessary for the functioning of the system, as shown later.

The last generation of on orbit GPS satellites is the block IIF, the first launch was in May 2010, and at the moment of this work only three space vehicle of block IIF were launched. These satellite are functionally equivalent to the IIR/IIR-M satellites and pave the way towards operational M-code and L2C signal. The improvement introduced by this satellite generation involves converting the GPS cesium clocks from analog to digital. Block IIF also adds a new separate signal for civilian use, designated L5 at the frequency of 1176.45 MHz. These two codes mark in the facts the transition to the GPS III era.

2.2.3 The future GPS III

GPS III block, will give new navigation warfare (NAVWAR) capabilities to shut off GPS service to a limited geographical location while providing GPS to US and allied forces. The new system will offer significant improvements in navigation capabilities by improving interoperability and jam resistance. The modernisation will be characterised by new navigation signal:

- The M-code is transmitted in the same L1 and L2 frequencies already used by the previous military code, the P(Y) code. It is designed to be autonomous, meaning that users can calculate their positions using only the M-code signal. It will be broadcasted from a high-gain directional antenna, in addition to a wide angle (full Earth) antenna, with the consequent possibility to highly improve at a very selective local scale the positioning performances, doubling the signals detected from the satellites. The IIF block satellite payload is not equipped with the the high gain antenna.

- The civilian L2 (L2C) signal an easy-to-track signal, providing improved accuracy of navigation, and acting as a redundant signal in case of localised interference. The immediate effect of having two civilian frequencies being transmitted from one satellite is the ability to directly measure, and therefore remove, the ionospheric delay error for that satellite. It is transmitted by all block IIR-M and

following satellites.

- The Safety of Life (L5) signal, a civilian-use signal, broadcasted on the L5 frequency (1176.45 MHz), implemented since the first GPS IIF launch (May 2010).

- The further L1 civilian-use (L1C) signal, to be broadcasted on the same L1 frequency (1575.42 MHz) that currently contains the C/A signal used by all current GPS users. The L1C will be available with first Block III launch, currently scheduled for 2014. The L1 signal will be interoperable with Galileo satellite navigation system.

- A further frequency (L4) will be available at 1379.913 MHz and will be under study for additional ionospheric corrections.

All these new signals will be transmitted with increased power and wider bandwidth and will ensure improvements on the services which they drive. The full operational constellation will consist of 32 satellites.

A new feature of the GPS III system will provide information about worldwide localisation of signals nuclear detonations, the Nuclear Detonation (NUDET) Detection System (NDS) payload. The NDS nuclear burst detectors will give capability to detect, locate, and report any nuclear detonations in the Earth's atmosphere, near space, or deep space in near real-time. The service will use optical, x-ray dosimeters, and ElectroMagnetic Pulse (EMP) detectors. The optical sensor measures atmospheric light signatures and their location. The EMP sensor measures atmospheric EMP phenomena to provide improved location accuracy and profile data. The X-ray sensor measures the exo-atmospheric X-ray spectrum while the dosimeter measures high energy proton and electron activities. GPS/NDS collects, processes, stores, and formats event data for transmission via an L3 communications link (1381.05 MHz). In addition, fixed site ground operators can command S-band memory read-outs of NDS data. The NDS payload, called Global Burst Detector (GBD), has been on board of satellites since Block IIA, with continuous evolution in subsequent satellite generations.

2.2.4 Satellite instruments

All satellites are equipped with dual solar arrays supplying over 400 W and chargeable NiCd batteries. A S-Band communication link (2227.5 MHz) is used for control and telemetry. An Ultra High Frequency (UHF) channel provides cross-links between spacecraft. A propulsion system is used for orbital correction. The payload includes two L-Band navigation signals at 1575.42 MHz (L1) and 1227.60 MHz (L2), locally generated starting from precise atomic oscillators. The satellite payload includes an antenna able to send a Right-Hand Circularly Polarised (RHCP) signal.

In the actual generation of satellite and status of the art of the instruments, the employment of latest technologies is able to give a growing resistance to jamming; the hardware is provided with reprogrammable processors that can receive software uploads, to improve the system versatility. The improvement involves converting the GPS cesium clocks from analog to digital processor-controlled cesium clocks continuously optimising their performance by adjusting internal parameters and compensating for environmental effects. They will also perform self-checkout diagnostics. The current analog clocks are optimised during manufacturing. Over the years, with the increasing age of components, the clock performance becomes less predictable and the output frequency of the clocks varies with changes in temperature or in the magnetic field. The digital cesium standard represents the next step in the evolution of frequency standards in space and promises to result in better accuracy for military and civilian GPS users.

2.2.5 Transmitter and receiver clock stability

As discussed before the most important factor for the correct mode of operation in the GPS system is the clock accuracy and the synchronisation capability between all the satellites. The clock reference and the clock offset parameters are also included in the navigation message, because the receiver clocks are not synchronised with satellite ones.

The core of the satellite payload is all along the atomic oscillator. Currently all satellites are provided by two Rubidium and two Cesium atomic clocks. One of these atomic standards is designated as primary and serves as timing reference on board of the space vehicle for navigation signal generation and transmission. These extremely accurate GPS atomic clocks must keep time to within a few nanoseconds a day and they must be synchronised each other. Despite the extreme precision, if the satellite clock drifts and offsets aren't take into account the resulting error in the positioning can become quite large. Furthermore the correction cannot be applied to the satellite clock, therefore drifts and clocks add up reaching values which must be known for the proper functioning of the system. The MCS of Colorado Springs collects all satellite data received by all the control segment stations; some advanced models and software allow to predict and compute future orbits and the parameters to refer the satellite clock to GPS reference time. The satellite clock correction is done by including in the navigation message the coefficient of a second order polynomial interpolation:

$$t_e = a_0 + a_1(t - t_r) + a_2(t - t_r)^2 \tag{2.1}$$

where: a_0, a_1, a_2 are the polynomial coefficients, t_e is the estimated (corrected) time, t_r is the reference time for the parameters computation, t is the uncorrected time. These parameters are uploaded to the satellites for realtime broadcasting, providing the real

time positioning and timing services. The stability and accuracy of a clock is often presented in terms of Allan variance ([15]):

$$\sigma_A^2 = \frac{1}{\tau^2}\left[\left\langle \phi(t+\tau)^2 \right\rangle \; - 2 \; \langle \phi(t+\tau)\cdot\phi(t) \rangle \; + \; \left\langle \phi(t)^2 \right\rangle\right] \tag{2.2}$$

where: $\sigma_A^2(\tau)$ is the Allan variance, τ is the averaging time [s], $\phi(t)$ is the clock signal phase [rad] at time t, $\langle\rangle$ indicates an infinite time average.
Following the Institute of Electrical and Electronic Engineers (IEEE) standard definitions ([16]), the Allan variance also called two-sample variance is computed using a finite series of samples. Being t_k the generic sampling time and using the definition:

$$\bar{y}_k = \frac{\phi(t_k+\tau) - \phi(t_k)}{\tau} \tag{2.3}$$

we write the Allan variance based on N samples:

$$\sigma_A^2(\tau) = \frac{1}{2(N-1)} \sum_{k=1}^{N-1} (\bar{y}_{k+1} - \bar{y}_k)^2 \tag{2.4}$$

Clock stability is usually expressed in terms of the Allan deviation, $\sigma_A(\tau)$ (the square root of the Allan variance). The most common oscillators available are quartz, rubidium cell, cesium beam, and hydrogen maser. Table 2.1 and Fig. 2.1 show the typical stability and thus accuracy through $\sigma_A(\tau)$ that can be expected from each oscillator.

	τ = **1 second**	τ = **1 day**	τ = **1 month**
Quartz	10^{-12}	10^{-9}	10^{-8}
Rubidium	10^{-11}	$10^{-12}/10^{-13}$	$10^{-11}/10^{-12}$
Cesium Beam	$10^{-10}/10^{-11}$	$10^{-13}/10^{-14}$	$10^{-13}/10^{-14}$
Hydrogen Maser	10^{-13}	$10^{-14}/10^{-15}$	10^{-13}

Table 2.1: Typical Oscillator Stabilities [17].

Quartz oscillators show an Allan deviation of 10^{-12} over short periods of about one second to one minute. This is comparable to a cesium beam and better than a rubidium cell over short periods. In the longer term, for instance a day or a month, quartz oscillators perform much worse than atomic standards. The stability of a hydrogen maser is about an order of magnitude better than the cesium beam for periods of up to one day. This very high short term accuracy and stability of quartz oscillators make them very usable in GPS receivers for positioning purposes requiring high receiver clock precision during the signal travel from satellite to ground station, and without need of synchronisation with GPS reference time.

2.3 GPS signal

The GPS signal is created on board of the space vehicle. Two frequency are included to allow for ionosphere delay computation and elimination. The carriers are in the L

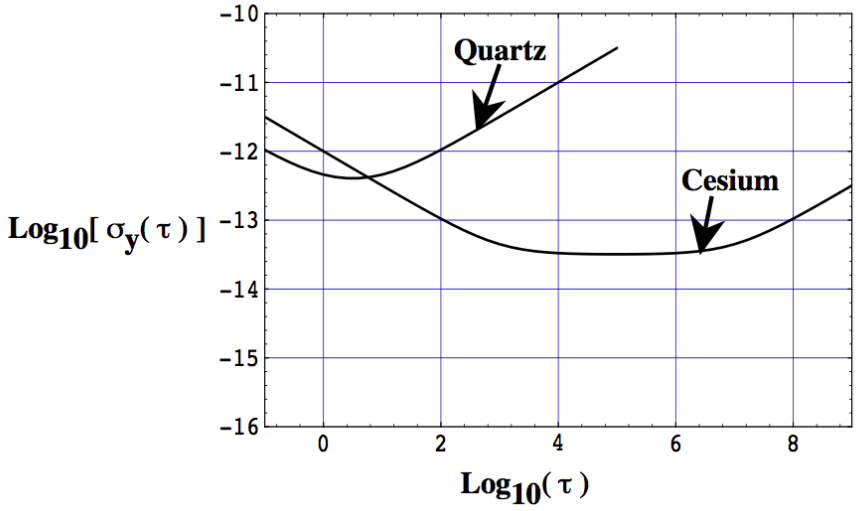


Figure 2.1: Typical Allan deviations of Cesium clocks and quartz oscillators, plotted as a function of averaging time τ. [18].

band (1 to 2 GHz) and originate from a common local oscillator fundamental frequency f_0 (nominally 10.23 MHz) by using Phase Locked Loop (PLL) frequency multipliers. Therefore the carrier signals (L_1 and L_2) are coherent with the same frequency clock:
$L_1 = f_0 \times 154 = 1575.42$ MHz
$L_2 = f_0 \times 120 = 1227.60$ MHz
The GPS carriers are modulated by signals which identifies univocally the transmitting satellite and provides information about its position. Three code types modulate the carrier signals (Fig. 2.2):

- The C/A code is the basis of the SPS. It is a bi-phase $(+1, -1)$ signal 1023 chips long. The chip rate is 1.023 Mchip s^{-1}, resulting in about 1 μs time length of each chip. The entire C/A code has the length of 1 ms. The spectrum of the signal (a square wave) is a *sinc* function with a null to null bandwidth of 2 MHz.

- The P code, or Y code when the encryption of A-S system is activated[1], is a bi-phase $(+1, -1)$ very long code, with a repetition time of 266 days (38 weeks). The chip rate is 10.23 Mchip s^{-1}. The whole P code is divided in 38 segments, each of them is 7 day long. Each segment is assigned to a satellite. Also this signal is a square wave and the spectrum bandwidth is 20 MHz.

- Finally the navigation data message (D) code with a bit rate of 50 bps, each bit is 20 ms long. It's a bi-phase $(+1, -1)$ signal. The navigation code include satellite ephemeris, time information, clock synchronisation parameters. All these

[1]The Anti-Spoofing mode is an encryption of the P code to create the Y code. The decryption of the Y code requires the use of an A-S module for each channel in the receiver and can be implemented only from authorised user.

parameters are required during the process of position determination.

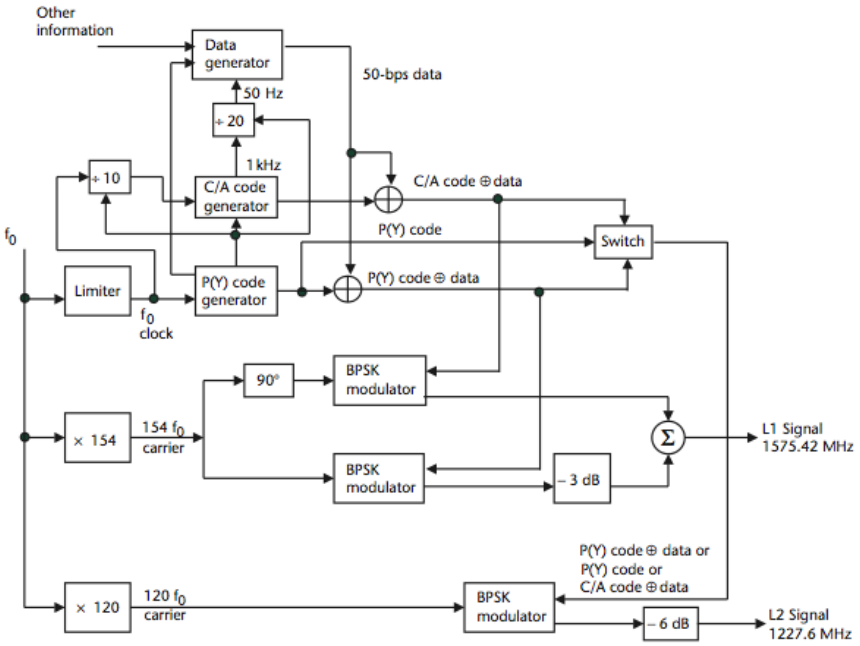


Figure 2.2: Scheme of the GPS satellite signal structure [19].

The transmitted signals in the two frequencies, S_{L_1} and S_{L_2} respectively, can be analytically written as:

$$\begin{aligned} S_{L_1} &= A_P P^k(t) D^k(t) \cos(2\pi f_1 t + \phi) + A_C C/A^k(t) D^k(t) \sin(2\pi f_1 t + \phi) \\ S_{L_2} &= A_P P^k(t) D^k(t) \cos(2\pi f_2 t + \phi) \end{aligned} \tag{2.5}$$

where:

A_P is the amplitude of the P code,

$P^k(t)$ is the precision code ($\pm$ 1) for the $k-th$ satellite,

$D^k(t)$ is the navigation code ($\pm$ 1) for the $k-th$ satellite,

A_C is the amplitude of the C/A code,

$C/A^k(t)$ is the coarse acquisition code ($\pm$ 1) for the $k-th$ satellite,

ϕ is the initial phase of the signal.

The GPS signal is a phase-modulated signal with modulating signals (C/A and P codes) varying between 0 and π; this type of phase modulation is referred to as Binary Phase Shift Keying (BPSK). C/A and P codes are in quadrature in order to allow the receiver to have a best detection of the signal.

The code signals can reach a minimum power level of -130 dBm and the spectrum is spread, so they cannot be detected from a spectrum analyser. The Spread Spectrum Modulation (SSM) is used to transmit the signal and the use of the Code Division Multiple Access (CDMA) allows to use the same band portion and the same centre frequency for all the signals transmission. The codes (C/A and P) relative to different

satellites are orthogonal: their cross correlation and their autocorrelation, computed for a non-zero shift delay, are very low. The autocorrelation function has a strong peak when the shift delay is zero, so that the receiver can correctly reconstruct the signal and retrieve all information (Fig 2.3). Often these codes are referred as Pseudo Random Noise (PRN), because the white noise has an autocorrelation function with the same properties. At the receiver level, the same code must be generated for each satellite in order to acquire any individual signal.

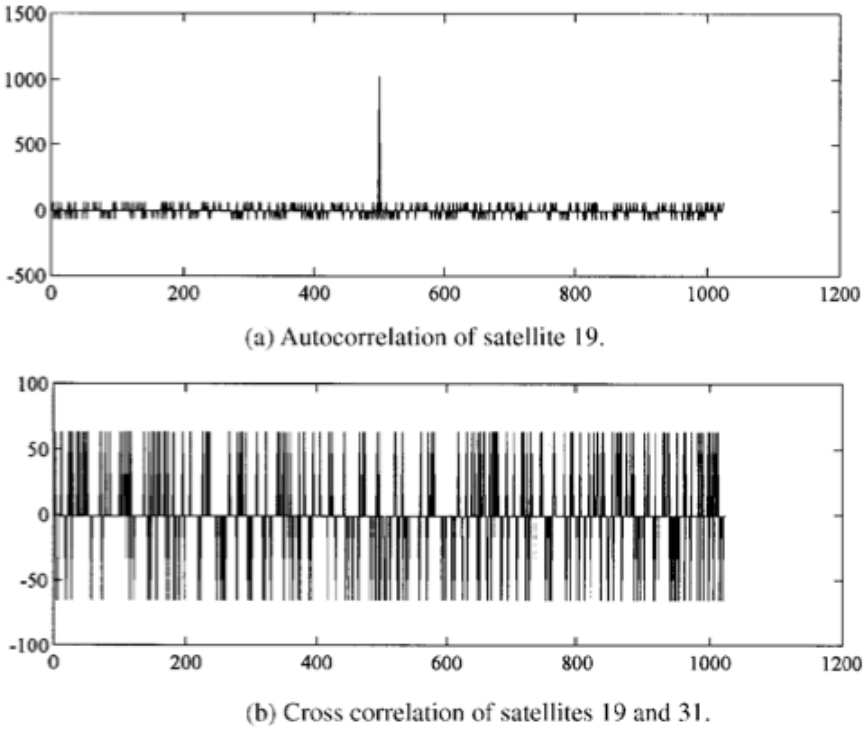


Figure 2.3: PRNs autocorrelation (a) and crosscorrelation (b) examples [20].

Also the receiver antenna must be RHCP to achieve the maximum efficiency in detecting signal.

2.3.1 Pseudorange measurements

The signal travel time is the time employed by the signal transmitted at time t_j^i by the i-th satellite[2] to reach the j-th receiver at time t_j; the receiver provides the travel time by computing the time shift needed to align the code received from satellite at time t_j with a replica generated at the same time in the receiver. The quantity obtained by multiplying this measurement by the GPS reference speed of light (299792458 m s^{-1}) is called pseudorange. The prefix "Pseudo" is due to the error sources that make different this quantity to the actual receiver satellite distance. The main error sources can be

[2]The transmission time is normally computed subtracting the pseudorange (travel time of the signal) from the instant of signal reception. For this reason it is related both to the i-th satellite sending time and to the j-th receiver reception instant ((t_j^i)).

analytically exploited in the following formula:

$$\begin{aligned}\rho^i_j(t_j,f) &= R^i_j(t_j) + c[\epsilon_{rj}(t_j) - \epsilon^i_s(t^i_j) + \Delta T^i_{Sat}(t^i_j,f) + \Delta T^i_{Trop\ j}(t_j) + \\ &+ \Delta T^i_{Ion\ j}(t_j,f) + \Delta T^i_{Rel}(t^i_j) + \Delta T^i_{Orbit}(t^i_j) + \\ &+ \Delta T_{Ric\ j}(t_j,f) + \Delta T^i_{Multi\ j}(t_j,f)] \end{aligned} \tag{2.6}$$

where:

$\rho^i_j(t_j,f)$ is the pseudorange measured by the j-th receiver at time t_j relative to the i-th satellite at the frequency f (L_1 or L_2);

$R^i_j(t_j)$ is the geometric distance between the i-th satellite position at time t^i_j of signal transmission and the j-th receiver position at time t_j;

$\epsilon_{rj}(t_j)$ is the j-th receiver clock shift due to the synchronisation error at time t_j;

$\epsilon^i_s(t^i_j)$ is the i-th satellite clock shift due to the synchronisation error at time t^i_j;

$\Delta T^i_{Sat}(t^i_j,f)$ is the satellite instrumental delay introduced by the i-th satellite at time t^i_j of signal transmission. It is dependent on the frequency of the signal. Different frequencies have different channel in the transmitter. It is referred to as T_{GD} (Group Delay);

$\Delta T^i_{Trop\ j}(t_j)$ is the delay introduced by the signal propagation into the troposphere, with respect to the free space propagation;

$\Delta T^i_{Ion\ j}(t_j,f)$ is the delay introduced by the signal propagation into the ionosphere, with respect to the free space propagation. It is frequency dependent;

$\Delta T^i_{Rel}(t^i_j)$ is the relativistic effect on the i-th satellite at time t^i_j.

$\Delta T^i_{Orbit}(t^i_j)$ is the delay caused by the error in the computation of the i-th satellite position at time t^i_j.

$\Delta T_{Ric\ j}(t_j,f)$ is the delay introduced by the receiver accuracy. It is primarily caused by the processing and propagation signal delay, but also by poor measurement accuracy. It is dependent on the frequency of the signal. Different frequencies have different channel in the receiver.

$\Delta T^i_{Multi\ j}(t_j,f)$ is a further delay introduced by multipath effect, caused by multiple signal reflections in the path from satellite to receiver antennas.

2.3.2 Carrier phase measurements

The signal carrier phase consist of an integer part and a fractional part, in terms of carrier cycles. Only the fractional part of the carrier phase can be directly measured by a receiver. Specific algorithms can be used in post-processing to estimate the integer part of carrier phase (ambiguity solutions). The j-th receiver measure the phase difference between the signal transmitted by the i-th satellite at time t_j^i, detected by the receiver at time t_j and the signal replica generated in the receiver. This phase measure is ambiguity affected, due to the unknown integer number of carrier cycles during the signal travel between satellite transmitter and receiver. Also carrier phase is affected by the same error sources as pseudorange.

$$\begin{aligned}
\Phi_j^i(t_j, f) &= \delta\phi_j^i(t_j, f) + N_j^i(t_j, f) = \\
&= \frac{f}{c} R_j^i(t_j) + f[\epsilon_{rj}(t_j) - \epsilon_s^i(t_j^i) + \Delta T_{Sat}^i(t_j^i, f) + \Delta T_{Trop\ j}^i(t_j) + \\
&- \Delta T_{Ion\ j}^i(t_j, f) + \Delta T_{Rel}^i(t_j^i) + \Delta T_{Orbit}^i(t_j^i) + \Delta T_{Ric\ j}(t_j, f) + \\
&+ \Delta T_{Multi\ j}^i(t_j, f)] \qquad (2.7)
\end{aligned}$$

Where the additional terms with reference to Eq. 2.6 are:

$\Phi_j^i(t_j, f)$ is the total phase including both integer and fractional part;

$\delta\phi_j^i(t_j, f)$ is the receiver measurement of the fractional part of carrier phase, hereafter referred as carrier phase measurement;

$N_j^i(t_j, f)$ is the integer number of carrier cycles during the signal propagation and it is the ambiguity of the measurement;

In Eq. 2.7 the ionosphere effect is opposite with respect to the ionosphere effect in Eq. 2.6, according with what was stated in § 1.4.

Next sections will briefly deal with the error sources listed in Eqs. 2.6 and 2.7.

2.3.3 Tropospheric delay

As explained in § 1.3 the path delay due to troposphere is caused due to air refractivity gradients. The air refractivity gradients in the troposphere are due to the dry air pressure, temperature, air moisture and condensed water in clouds or rain. The troposphere delay correction is made in most of cases using models of ZHD (e.g Eq. 1.34 express the Saastamoinen zenith delay model). The ZWD is difficult to model due to the high variability of water vapour distribution. However, the hydrostatic delay correction corrects approximately the 80% of the ZTD. Once the approximated ZHD is computed, the corresponding Slant Hydrostatic Delay (SHD) must be computed, by means of a geometric projection along the satellite receiver direction. This projection

can be made using a so called mapping function, a semi-empirical function which introduces the obliquity factor as a function of elevation of the satellite and of empirically derived parameters (see § 3.1 for more details). Different type of mapping functions are generally used for the slant wet delay and the slant hydrostatic delay respectively. Several mapping functions exist and have been studied and validated in literature. For some applications more accurate corrections can be reached using 3D meteorological output fields of numerical weather prediction models. This involves the tracing of the satellite receiver geometric path and the computation of the refraction index and the corresponding delay. The parameters for real time estimation of troposphere path delay are broadcasted by the satellites of the Space Based Augmentation System (SBAS) together with parameters for ionosphere delay estimation (§ 2.3.4) and satellite clock bias correction (§ 2.3.9).

2.3.4 Ionospheric delay

As explained in § 1.4.3, the refraction of the electromagnetic signals in the ionosphere acts as a slowing effect, and deviating the propagation direction (bending effect). As discussed in § 1.5 the ionospheric delay is the addition of both effects, resulting in a delay on the signal travel time resulting in a term of tens of meters in the pseudorange value. The ionospheric delay is strongly dependent on the portion of the ionosphere crossed by the signal, which increases at lower elevations of the receiver-satellite direction. The dispersion characteristic of the ionosphere produces opposite effects on the group and phase velocity of the signal, and consequently on the pseudorange and carrier phase measurements. From Eq. 1.62, the difference ($\Delta\tau$) between group delay (or phase advance) in the two different frequencies (f_{L1},f_{L2}) is:

$$\begin{aligned}\Delta\tau &= B\left[\frac{1}{f_{L2}^2}-\frac{1}{f_{L1}^2}\right]=\frac{B}{f_{L2}^2}\left[\frac{f_{L1}^2-f_{L2}^2}{f_{L1}^2}\right]= \\ &= \Delta T_{Ion}(f_{L2})\left[1-\frac{f_{L2}^2}{f_{L1}^2}\right]= \\ &= \Delta T_{Ion}(f_{L1})\left[\frac{f_{L1}^2}{f_{L2}^2}-1\right] \qquad (2.8)\end{aligned}$$

Considering P_1 and P_2 the pseudorange measurements in the *precision-code*, C_1 e C_2 the pseudorange measurement in the *C/A-code* and L_1 e L_2 the carrier phase measurements, by assuming negligible the variation with the frequency of all terms present (different from the ionospheric one) in Eq. 2.6 and starting from Eq. 2.8:

$$\begin{aligned}\Delta T_{Ion}^{P_2}(f_{L2}) &= \frac{[P_2-P_1]}{c}\frac{f_{L1}^2}{f_{L1}^2-f_{L2}^2} \\ \Delta T_{Ion}^{P_1}(f_{L1}) &= \frac{[P_2-P_1]}{c}\frac{f_{L2}^2}{f_{L1}^2-f_{L2}^2}\end{aligned}$$

$$\begin{aligned} \Delta T_{Ion}^{C_2}(f_{L2}) &= \frac{[C_2 - C_1]}{c} \frac{f_{L1}^2}{f_{L1}^2 - f_{L2}^2} \\ \Delta T_{Ion}^{C_2}(f_{L1}) &= \frac{[C_2 - C_1]}{c} \frac{f_{L2}^2}{f_{L1}^2 - f_{L2}^2} \end{aligned} \tag{2.9}$$

Unfortunately the frequency dependent terms introduced by the satellite and receiver hardware are not negligible and must be modeled and taken into account (as shown in § 2.3.6) to accurately estimate the ionospheric effect.

Furthermore it is not possible to use only carrier phase measurements in two different frequencies for computing the ionosphere delay, because of the ambiguities. Some ancillary data (such as pseudorange measurements) must be used to solve these ambiguities.

The introduction of triple frequency systems as future GPS III and Galileo, will allows to improve the accuracy in the estimation of ionosphere effects, by using a second order approximation (see Eq. 1.64).

2.3.5 Relativistic effects

For measuring and determining the orbits of satellites, the GPS system uses an inertial cartesian coordinate system, the Earth Centreed Inertial (ECI) reference system in which the origin is at the centre of mass of the Earth, the x-y plane is coincident with the Earth's equatorial plane and the x, y axis oriented along determined positions over the celestial sphere; the z axis is normal to the x-y plane in the north direction. Because of the large motion velocities and near circular orbits of the GPS satellites, the non-negligible gravitational potential difference between the satellite and the users, as well as the rotation of the Earth, the relativistic effects have to be taken into account.

The nominal frequency of 10.23 MHz at the Earth observer is created by a proper correction of the atomic oscillator frequency, in order to take into account for the average relativistic effect caused by the satellite orbital motion and gravitational potential difference. The averaged relativistic effect on the signal frequency can be approximated by:

$$-\frac{f_0 - f'}{f'} = \frac{1}{2}\left(\frac{v}{c}\right)^2 + \frac{\Delta U}{c^2} \tag{2.10}$$

Where f' is the working frequency, f_0 is the nominal frequency (10.23 MHz), v is the satellite speed, c is the light speed and ΔU is the gravitational potential difference between satellite and receiver. The f' working frequency must be chosen and set to verify the Eq. 2.10. To correct for average relativistic effects, the parameters are taken as mean value; the reference frequency on board of space vehicle is decreased by a factor 4.567×10^{-3} Hz in the design phase, before the satellite launch; therefore the core frequency of GPS satellites is $f' = 10.23 \times 10^6 - 4.567 \times 10^{-3} = 10.229999995433$ MHz. At an Earth-based GPS receiver the frequency result to be the nominal frequency

($f' = f_0$).

Eq. 2.10 considers a circular orbit for all GPS satellites. To take into account for the orbit eccentricity (that has consequences in the gravitational potential differences) a correction must be applied to the pseudorange and carrier phase relationship (2.6) and (2.7):

$$\Delta T_{Rel} = \frac{2}{c^2}\sqrt{\mu a}\ e\ \sin E \tag{2.11}$$

where $\mu = 3986005 \times 10^8$ m^3 s^{-2} is the product of the universal gravitational constant and the Earth mass, a is the semimajor axis, e is the orbit eccentricity, E is the eccentric anomaly (to be computed using the Kepler equation, for example as shown in § 2.5.3.1). The value of a and e are part of the broadcasted ephemeris.

In any case some relativistic effects caused by the orbit oscillations, Earth gravitational field perturbation etc, persist. Furthermore the satellite and receiver relative motions can produce a Doppler effect approximately within ±5 kHz ([20]). These effects can be modeled and precisely estimated in order to give negligible contribution to the overall error in positioning or, in the case of our interest, in the tropospheric delay evaluation. Here follows some examples:

The general relativity effects of the signal can be represented by the Holdridge (1967) model ([10]):

$$\delta\rho_{Rel} = \frac{2\mu}{c^2}\ln\frac{\rho_i + \rho^j + \rho_i^j}{\rho_i + \rho^j - \rho_i^j} \tag{2.12}$$

where ρ_i is the geocentric position of the i-th receiver, ρ^j is the geocentric position of the j-th satellite, ρ_i^j is the geocentric distance between the i-th receiver the j-th satellite. $\delta\rho_{Rel}$ can reach a maximum value of about 2 cm.
The correction for the acceleration of the Earth satellite, following the model of McCarthy 1996, is([10]):

$$\Delta\mathbf{a} = \frac{\mu}{c^2 r^3}\left\{\left[4\frac{\mu}{r} - v^2\right]\mathbf{r} + 4(\mathbf{r}\cdot\mathbf{v})\mathbf{v}\right\} \tag{2.13}$$

where $\mathbf{r}$ is the geocentric satellite position vector, $\mathbf{v}$ is the geocentric satellite velocity vector and $\mathbf{a}$ is the geocentric satellite acceleration vector.

In addition to all the previous terms, there is a further factor due to relative motions of satellite and receiver. A receiver can be generally in motion on the Earth's surface. For our purposes, however, only receiver with fixed position over the Earth surface will be taken into account. Thus the only effect that we consider is known as Sagnac effect, which is due to the Earth rotation during the signal travel time. This effect will be described in § 2.3.8.

2.3.6 Satellite and receiver instrumental delays and differential code biases

There are different instrumental biases for different codes and different frequencies. The available pseudorange observables are C1 (C/A code pseudorange on the f_1 frequency), P1 (P code pseudorange on the f_1 frequency), P2 (P code pseudorange on the f_2 frequency); for each of them, we can identify the biases:

- B_{C1} : bias on the C1 observable
- B_{P1} : bias on the P1 observable
- B_{P2} : bias on the P2 observable

These are instrumental biases, so there are two biases for each observable: one produced by the satellite and the other produced by the receiver. Separately for the satellite and for the receiver we have to consider the discrepancies between true (superscript T), corrected from the biases and measured (superscript M), affected by biases, values:

$$\begin{aligned} C_1^M &= C_1^T + B_{C1} \\ P_1^M &= P_1^T + B_{P1}; \\ P_2^M &= P_2^T + B_{P2}; \end{aligned} \tag{2.14}$$

The absolute value of single biases are generally not directly retrievable because of the variability of instrumentation biases caused by the changes of satellite instrument working conditions (such as temperature) and status (wear) and because the instrumental biases of receiver and satellite adds up into measurements, so it's common the use of biases combinations, commonly referred to as Differential Code Biases (DCB):

$$\begin{aligned} B_{P1-P2} &= B_{P1} - B_{P2} \\ B_{P1-C1} &= B_{P1} - B_{C1} \end{aligned} \tag{2.15}$$

These combination cannot directly be applied for a single observable correction, but to a combination of observables. For example considering the availability of observables C1 and P2 from a receiver the known approximated model of pseudoranges (see Eq. 2.6), by neglecting the time (transmission and reception) dependence, can be written as:

$$\begin{aligned} C1_j^i &= R_j^i + c\left[\epsilon_{rj} - \epsilon_s^i + \Delta T_{Trop} + \Delta T_{Ion}(f_1) + \Delta T_{Rel}^i + \Delta T_{Orbit}^i\right] + B_{C1} \\ P1_j^i &= R_j^i + c\left[\epsilon_{rj} - \epsilon_s^i + \Delta T_{Trop} + \Delta T_{Ion}(f_1) + \Delta T_{Rel}^i + \Delta T_{Orbit}^i\right] + B_{P1} \\ P2_j^i &= R_j^i + c\left[\epsilon_{rj} - \epsilon_s^i + \Delta T_{Trop} + \Delta T_{Ion}(f_1) + \Delta T_{Rel}^i + \Delta T_{Orbit}^i\right] + B_{P2} \end{aligned} \tag{2.16}$$

where

$B_{C1} = \Delta T_{Sat}(t_j^i, C_1) + \Delta T_{Ric}(t_j, C_1)$ is the bias introduced on the C1 observable by the instrumental delays of the satellite and the receiver;

$B_{P1} = \Delta T_{Sat}(t_j^i, P_1) + \Delta T_{Ric}(t_j, P_1)$ is the bias introduced on the P1 observable by the instrumental delays of the satellite and the receiver;

$B_{P2} = \Delta T_{Sat}(t_j^i, P_2) + \Delta T_{Ric}(t_j, P_2)$ is the bias introduced on the P2 observable by the instrumental delays of the satellite and the receiver.

The computation of the delay introduced by the ionosphere involves the combination of the dual frequency observations, as shown in Eq. 2.9. Therefore, in the computation of the ionospheric delay, we have to compute the quantity:

$$\begin{aligned} P_2 - C_1 &= \Delta T_{Ion}(f2) - \Delta T_{Ion}(f1) + B_{P2} - B_{C1} = \\ &= \Delta T_{Ion}(f2) - \Delta T_{Ion}(f1) - B_{P1} + B_{P2} + B_{P1} - B_{C1} = \\ &= \Delta T_{Ion}(f2) - \Delta T_{Ion}(f1) - B_{P1-P2} + B_{P1-C1} \end{aligned} \tag{2.17}$$

The quantities B_{P1-P2} and B_{P1-C1} are DCBs and must be know for the complete correction of the ionospheric delay. Code biases are instrument dependent bias, so there are different code bias for satellite and receiver instruments. Also differential code biases are different for satellites and receivers; methods to estimate with acceptable accuracy the satellite ones have been investigated from several years from the GPS community. The values of satellites code biases B^i_{P1-P2} are referred to as τ^i_{GD} in the Interface Control Document ([21]), are computed by ground control stations and transmitted by satellites in the navigation message. The values of B^i_{P1-C1} can be corrected by applying tables that are constantly updated. With reference to the receiver biases the user must compute them during data processing.

The relationship between B^i_{P1-P2} and τ^i_{GD} is:

$$\tau^i_{GD} = \frac{1}{c(1-\gamma)} B^i_{P1-P2} \simeq -1.55 B^i_{P1-P2} \tag{2.18}$$

Being: $\gamma = \frac{f_1^2}{f_2^2} \simeq 1.65$ Equation 2.17 will become:

$$\begin{aligned} P_2 - C_1 &= \Delta T_I(f2) - \Delta T_I(f1) + c(\gamma - 1)\tau^i_{GD} + \\ &+ B^i_{P1-C1} - B_{jP1-P2} + B_{jP1-C1} \end{aligned} \tag{2.19}$$

In order to correctly compute the ionospheric delay, we have to compute all the differential code biases not given in the broadcasted message (B^i_{P1-C1}; B_{jP1-P2}; B_{jP1-C1}). When all these contribution are known the ionospheric delay is:

$$\begin{aligned} \Delta T_I(f2) &= \frac{\gamma}{\gamma - 1}\left[P_2 - C_1 - c(\gamma-1)\tau^i_{GD} - B^i_{P1-C1} + B_{jP1-P2} - B_{jP1-C1}\right] \\ \Delta T_I(f1) &= \frac{1}{\gamma - 1}\left[P_2 - C_1 - c(\gamma-1)\tau^i_{GD} - B^i_{P1-C1} + B_{jP1-P2} - B_{jP1-C1}\right] \end{aligned} \tag{2.20}$$

As mentioned before the quantity B^i_{P1-C1} can be computed using methods of proven accuracy and well known in the GPS community. Some software are freely provided for the correction of this satellite-based bias.

On the contrary for the computation of the receiver-based DCBs there are no standard methods. In order to make an approximate estimation for DCBs we have compared measures from some GPS station in Tuscany with theoretical delays obtained from a standard model of the ionosphere, the Klobuchar model ([12]), which can estimate and remove most of the ionospheric effects and is used as reference model by the GPS community. The Klobuchar model (see § 1.4.3) allows to compute the TEC by means of an algorithm that parameterise the sun activity. In order to correct for seasonal and sun spot number changes, the algorithm uses different sets of eight coefficient depending upon the period of the year and the average solar flux. Better estimations are obtained during night periods, when part of the ionisation reach more stable values (see § 1.2.1). Fig. 2.4 shows a plot of the ionosphere delay computed by using Eq. 2.20 taking into account the satellite-based instrumental delays but not the receiver ones and the Klobuchar model-based ionospheric delay (Eq. 1.65), computed for the same period.

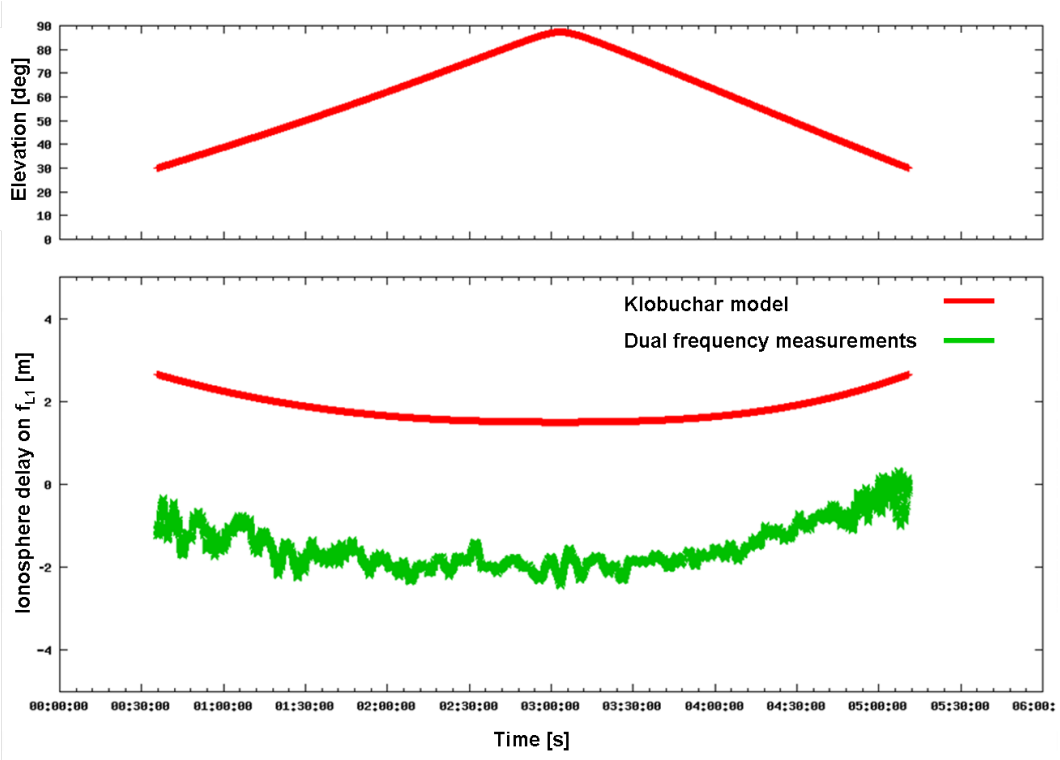


Figure 2.4: Top panel: elevation angle as a function of time. Bottom panel: Klobuchar model ionospheric delay (red) and dual frequency ionospheric delay affected by biases (green). Station mass (Massa, Tuscany). Data are relative to 2010 28^{th} February.

The ionospheric delay based on two frequency computation is affected by the receiver DCBs. The Klobuchar model during the minimum value of TEC relative to

minimum values of sun activities (i.e during night time near the sunset and for high satellite elevation angles) can be used to estimate the value of the receiver DCB (comprehensive of B_{jP1-P2} and B_{jP1-C1}), as a free parameter whose value minimises the distances between two curve minima.

2.3.7 Orbit parameters and errors

Ephemeris data, broadcasted in the satellite navigation message, give the satellite positions as a function of time. They are predicted starting from the current satellite positions estimated by the ground control stations. Typically, overlapping intervals of 4-hour GPS data are used by the operational control system to predict satellite orbital elements for each 1-hour period. Of course, both the current satellite position estimation and the future satellite position prediction can be error affected. Ephemeris errors vary from 2 m to 5 m, and can reach up to 50 m under selective availability. The ephemeris error is usually decomposed into components along three orthogonal directions defined by the satellite orbit: radial, along-track and cross-track (see Fig. 2.5).

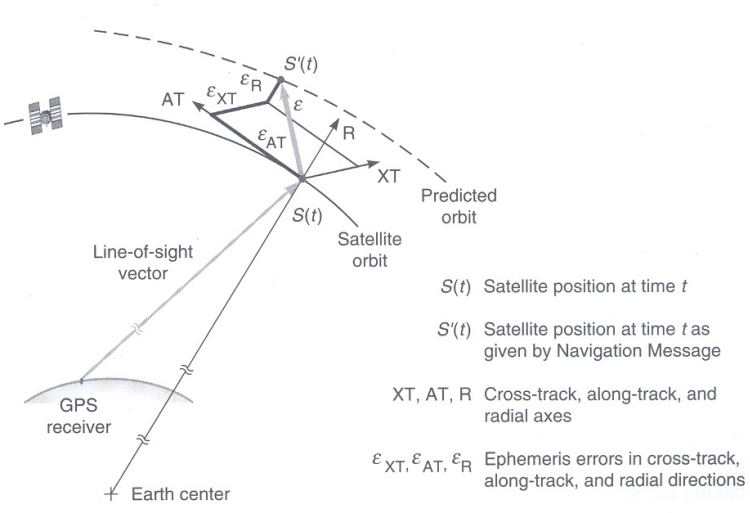


Figure 2.5: Ephemeris error components. ([22]).

Errors sources impacts in the pseudorange and carrier phase measurements, by their projection in the receiver-satellite vector direction (Line-of-sight). The projection of the satellite position error in the line-of-sight direction depends primary on the radial component and secondary on cross-track and along-track components. Fortunately the radial error component in the estimation and prediction of satellite position is smaller than others by one order of magnitude. Due to the fact that this errors originate from a prediction, they increase with the lapsing time from the instant of the last upload of the orbital parameters by the control station.

The increasing need of precise ephemeris in quasi-real time for an increasing number of application has concentrated the efforts of the scientific community to develop

products more and more precise with respect to ephemeris broadcasted in the navigation message. Several institutions, e.g., the International GNSS Service, formerly the International GPS Service for Geodynamics (IGS), the U.S. National Geodetic Survey (NGS), and Geomatics Canada, have developed postmission precise orbital service. Precise ephemeris data are based on GPS data collected at a global GPS network coordinated by the IGS. Precise ephemeris data contain very accurate parameter to correct the space vehicle clock offset and drift and to retrieve the satellite position at transmission time. At the present time, precise ephemeris data are available to users in real time Ultra-Rapid (see [23]) with an accuracy of 5 cm for the ephemeris prediction.

The use of differential techniques combining measurements from different receivers positions cannot totally remove the ephemeris error. Only if the distance between receivers is short (short baseline) the error range due to ephemeris can be considered the same and can be easily removed by differencing the observations.

2.3.8 Earth rotation effect: the Sagnac effect

In the geometric distance computation the satellite and receiver positions must be referred to the same reference system. For positioning purposes the reference coordinate system is the Earth Centreed Earth Fixed (ECEF) system (§ 2.5.4), a cartesian reference frame with the X-Y plane coincident with the equatorial plane and the X direction oriented toward the Greenwich meridian. Z axis is oriented to the north polar axis and the reference frame origin is the centre of mass of the Earth. With reference to Fig. 2.6 the satellite position computed from ephemeris is referred to the transmission time (superscript T). The receiver position is referred to the reception time (superscript R). For a fixed receiver in the ECEF reference system the position is constant in time. Since the ECEF reference system rotates with the Earth during the signal travel time the receiver position and the satellite position are not referred to the same reference system. This effect is known as the Sagnac effect. To take into account for it, the positions of receiver and satellite must be referred to the same reference frame (i.e ECEF at transmission time or ECEF at reception time).

Now let us show the computation of the position of satellite (at transmission time) in the ECEF reference frame at reception time. As shown in Fig. 2.6, during the signal travel time (Δt), the geometric distance changes for the Earth rotation effect. The Earth rotation ($\dot{\omega_e}$ is the rotation angular speed) causes the reference system rotation of an angle $\dot{\omega_e}\Delta t$. The relationship to have the coordinates of satellite position at the transmission time referred to the ECEF reference system at reception time are:

$$\begin{cases} X^R = X^T \cos(\dot{\omega_e}\Delta t) + Y^T \sin(\dot{\omega_e}\Delta t) \\ Y^R = -X^T \sin(\dot{\omega_e}\Delta t) + Y^T \cos(\dot{\omega_e}\Delta t) \\ Z^R = Z^T \end{cases} \tag{2.21}$$

This simply solves the Sagnac effect for GNSS signal processing. The Sagnac effect

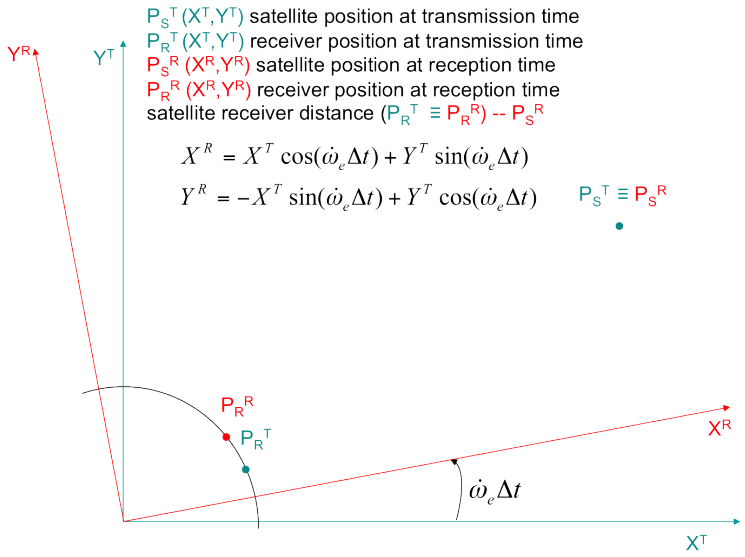


Figure 2.6: Effects of Earth rotation: reference system rotation.

must be considered in the procedures for synchronisation of clocks all over the globe, including satellite ones.

2.3.9 Satellite clock offsets and drifts

The GPS system uses atomic clock with Cesium or Rubidium oscillator. They have a nominal precision of about 1 part of 10^{13} (see the Allan deviation in Fig. 2.1). In a day (86400 s) the satellite clock offset can reach the order of 10^{-7} s, that multiplied by the speed of light gives an equivalent distance error of about 2.5 m. If the prediction of clock behavior could be done just with this accuracy the resulting precision would be of this order. Using updates of clock correction prediction every 12 hours we can assume an error of $1 - 2$ m. The clock predicted parameters are phase bias, frequency bias, and frequency drift rate. They are uploaded to the satellites and broadcasted into the navigation message. As can be easily inferred, the ephemeris and satellite clock errors are closely related. As mentioned for Ephemeris, the effort of scientific community is in the direction of the implementation of models for the prediction of clock correction parameters with increasing accuracy. In real time Ultra-Rapid ephemeris there are also clock correction parameters predicted with a nominal accuracy of 1.5 ns, i.e 0.45 m.

The residual satellite clock errors are the same in all measurement involving the same satellite, including P-code, C/A code and carrier phase, and can be easily removed by using differential techniques.

2.3.10 Receiver clock errors

GPS receivers commonly use inexpensive crystal clocks, which are much less accurate than the satellite clocks (see the Allan deviation in Fig. 2.1). The quartz receiver clock is more stable than satellite atomic ones, for instantaneous observations. Therefore the

quartz clocks are very suited for positioning purposes that require high stability for very short periods (tens of nanoseconds) that go from the transmission instant to the reception and measurement instant.

The receiver clock offset including receiver and satellite instrumental biases, can be considered as a constant for all observations regarding different satellites on view at the same instant. As will be shown in § 2.5, in the position problem the receiver clock error is a further unknown in addition to three coordinates unknowns that can be computed (estimated) by solving a system of equations, once the other error sources are neglected or separately estimated.

The receiver error can be larger than satellite clock errors; however it can be totally removed by making differences between observation relative to different satellites. Clock stability depends critically on the type of oscillator in use. A crystal oscillator is an electronic oscillator circuit that uses the mechanical resonance of a vibrating crystal of piezoelectric material to create an electrical signal with a very precise frequency. It is thus possible to obtain stabilised frequencies, that, however are influenced by temperature. In order to cope with temperature instabilities various forms of compensation are used, from analog compensation (TCXO) to stabilisation of the temperature through a crystal oven (OCXO)[3]. OCXO are the most stable oscillators, as they are actively stabilised in temperature. TCXO on the contrary has a sort of feedback mechanism that realigns frequencies when, due to temperature variations, they jump beyond predetermined thresholds. In Fig. 2.7 there are two example measurements of GPS signals we have processed from two receivers of the GPS network of the Regione Toscana, mounting these two different kinds of oscillators. We have used differential measurements with a time lag of $\Delta t = 1$ s of pseudorange ρ and satellite distance from the receiver R, in order to eliminate (at a first order) slow varying contributes (due to atmosphere, antennas, biases, etc.). We have also computed a ionospheric-free signal and from this a "total" bias B minus the tropospheric delay, $c\,\tau_{Trop}$, that we expected to possess a short term stability. B is the sum of three terms, the first consists of known parameters, namely the interfrequency bias B_{2-1} and the ratio of the square frequencies, γ; the second is the clock bias, b_c, and the third is the signal processing bias, $b(f_2)$, that are unknown. We can easily observe the dramatic difference in stability of the two oscillators, and the large realignment jumps of 100 ns (equivalent to about 30 m) for TCXO measurements, that brings to a very unstable bias, much greater than precision constraints for many applications (our included). This means that single TCXO measures are essentially unusable, and it is necessary to process reasonable ensembles of measurements in order to obtain enough precise results at least for positioning.

[3]TCXO stays for Temperature Controlled Crystal Oscillator; OCXO stays for Oven Controlled Crystal Oscillator (being "XO" an old acronym for "crystal oscillator").

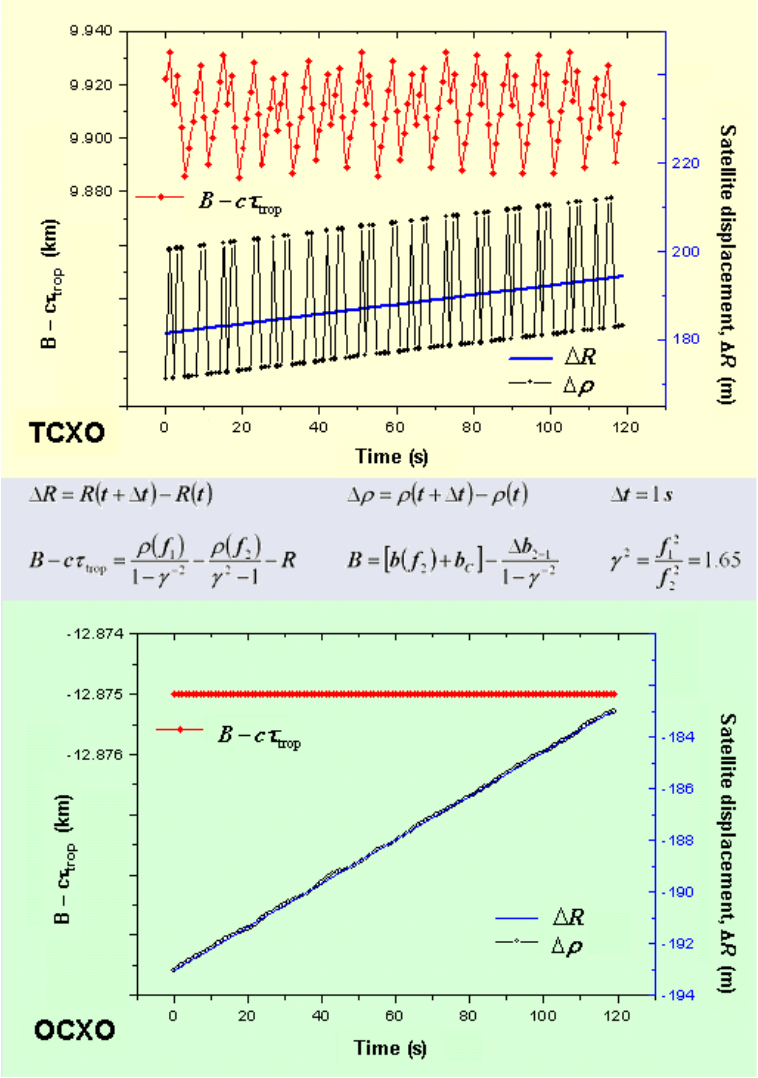


Figure 2.7: Stability of TCXO and OCXO clock crystals (top and bottom panels respectively). R is the satellite distance from the receiver, ρ the pseudorange, B is the "total" bias, containing the interfrequency bias B_{2-1}, the clock bias b_c, and the signal processing bias $b(f_2)$; $c\,\tau_{Trop}$ is the tropospheric delay. Figures are from two different receivers of the GPS network in Tuscany.

2.3.11 Multipath

If the GPS signal reaches the receiver after reflections from some obstacles, or from the Earth surface, we can speak of "multipath". The presence of indirect signals that arrive later than the direct ones and interfere with them, causes waveform distortion but also phase distortion. Multipath errors affect both carrier phase and pseudorange measurements; but the effects in the carrier phase measurement is limited to a quarter of a cycle and in the pseudorange measurements can reach several tens of meters (in the C/A code measurements). The problem is that the error on pseudorange measurements can degrade the ambiguity resolution process required for using carrier phase measurements. It doesn't exist a fully efficient mitigation technique of the multipath effects, but a series of commonly used methods and best practices. The use of differential GPS cannot reduce the multipath effects. The best ways to mitigate multipath effects are:

- to choose an installation site with a limited number of near potential reflecting object;
- to use a groundplane antenna (i.e. choke ring) able to attenuate signal reflected from the ground;
- to use a directive antenna array with many directive patterns simultaneously operative and able to adapt to the changing satellites geometry;
- to use a hardware (antenna gain) or software (signal processing) selected "cutoff" angle (i.e. minimum desired elevation angle) to eliminate measurements with low satellites elevation angles which can have higher multipath effects;
- to use a polarised antenna able to match the polarisation of the GPS signal (i.e., right hand) but mitigating in this case only single reflection or odd number of reflection effects.
- to observe for long time intervals the signal, in order to detect sudden change caused by multipath (only for fixed receiver);
- to use techniques and technologies based on narrow correlation of the receiver that allow to automatically suppress signals delayed by more than 1.5 chips (a chip corresponds to 1 pulse: about 1 μs for the C/A-code and about 100 ns for the P-code);
- to analyse the shape of the correlation function.

Other methods exist but these are the main used. For our purpose it is sufficient to mention that choke ring antenna have been chosen for all the used fixed station and they have been installed over roofs of buildings and generally far enough from sources of reflections.

2.3.12 Additional error sources

Some other error sources can degrade the receiver measurements. They normally generate errors negligible compared to those listed above. However, by accounting for them, it is possible to achieve greater accuracies both in terms of positioning and in terms of GNSS meteorology, which is the main objective of this work.

- Cycle sleep is a phenomenon due to a loss of satellite signal track, affecting carrier phase measurements. Some causes can be the presence of obstacles along the signal ray, radio interference and strong ionospheric activity. The cycle sleeps can be identified and corrected in several ways, such as analysing the difference between two double difference (i.e "triple differences", see § 2.3.13.2 for more details).

- The antenna phase centre generally doesn't coincide with the geometrical centre of the antenna. As the GNSS signal is circularly polarised the antenna phase centre varies in function of the elevation and the azimuth of the satellite signal ray. This error depends on the antenna type and typical values are of the order of a few centimeters. The effects can be mitigated if the attitude and azimuth of the antenna are known (for fixed receivers). The point in the antenna that is used to determine the offset in relation to the geodetic point above which the antenna is installed, is the Antenna Reference Point (ARP) and this represent an offset value. The Phase Centre Variations (PCV) can vary in function of the signal. The NGS ([24]) and the IGS ([25]) provide and update absolute and relative antenna calibration files, valid for a wide range of GNSS antennas.

- The antenna coordinates must be corrected also for some additional effects: the solid Earth tides errors (up to several decimeters) are caused by the gravitational field perturbation in the Earth Sun Moon system. Their periodicity can be derived from motion of celestial bodies and the effects on position error can be computed and corrected with good approximation. Minor effects originate from ocean loading (errors up to some millimeters), a deformation of the distribution of ocean water and coastal land due to ocean tides. Polar motion is a further error source (up to a few centimeters) affecting the determination of the receiver and satellite position in the reference frame. Finally also the Tectonic Plate Motion must be known and corrected in the position accuracy determination. Some geophysical models can be used; very good accuracy is provided by GNSS geodetic networks used to detect movements of the Earth's crust.

- The selective availability (disabled in May 2000) were a further error source, due to two artificial effects: the first called δ error is a degradation of the satellite clock frequency, the second is the ϵ error and is an additional artificial orbital error. This type degradation affects both carrier phase and pseudorange measurement.

2.3.13 GPS Observables

In this chapter the main types of "observables" used in the GPS community are briefly analysed and described. Some observables can be obtained by using a single receiver and relative to a single satellite, some other by using two receiver and two satellite making differences.

2.3.13.1 Single receiver observables

The basic observables are obtained directly by receiver measurements. With reference to Eq. 2.6 and Eq. 2.7, neglecting the superscript and subscript referred to the

considered receiver and satellites:

$$\begin{aligned} P &= \rho \\ L &= \lambda\, \delta\phi \end{aligned} \tag{2.22}$$

1. P_1, L_1 the pseudorange and phase measurements in the first frequency ($f_1 = 1575.42$ MHz) respectively

2. P_2, L_2 the pseudorange and phase measurements in the second frequency ($f_2 = 1227.6$ MHz) respectively

3. P_{IF}, L_{IF} the pseudorange and phase linear combinations respectively, called *Ionosphere free LC* because this linear combination nearly completely eliminates the ionospheric refraction effects:

$$\begin{aligned} P_{IF} &= \frac{1}{f_1^2 - f_2^2}(f_1^2 P_1 - f_2^2 P_2) \\ L_{IF} &= \frac{1}{f_1^2 - f_2^2}(f_1^2 L_1 - f_2^2 L_2) \end{aligned} \tag{2.23}$$

4. P_{GF}, L_{GF} the pseudorange and phase linear combinations respectively, called *Geometry free LC* because their linear combination cancel the frequency independent part of the measurement, including the geometric distance, leaving only the ionospheric effects and the instrumental constants (multipath if present, instrumental biases, and other observational noises):

$$\begin{aligned} P_{GF} &= P_1 - P_2 \\ L_{GF} &= L_1 - L_2 \end{aligned} \tag{2.24}$$

5. P_{NW} and P_{NW} the pseudorange and phase linear combinations respectively, called *Narrow Lane LC*. It has the lowest noise level of all used linear combinations, and consequently yields to the best results. It is characterised by $\lambda_{NW} = 10.7$ cm and an ambiguity difficult to resolve:

$$\begin{aligned} P_{NW} &= \frac{f_1 P_1 + f_2 P_2}{f_1 + f_2} \\ P_{NW} &= \frac{f_1 L_1 + f_2 L_2}{f_1 + f_2} \end{aligned} \tag{2.25}$$

6. P_{WL} and P_{WL} the pseudorange and phase linear combinations respectively, called *Wide Lane LC* that giving a large wavelength ($\lambda_{WL} = 86.2$ cm) four times bigger than f_{L1}, is useful for cycle slip detection or ambiguity resolution. The disadvantage is the high noise level of this signal:

$$\begin{aligned} P_{WL} &= \frac{f_1 P_1 - f_2 P_2}{f_1 - f_2} \\ L_{WL} &= \frac{f_1 L_1 - f_2 L_2}{f_1 - f_2} \end{aligned} \tag{2.26}$$

7. L_{MW} is a linear combination of both carrier and pseudoranges observables, called *Melbourne Wubbena LC*, that's exactly the difference between L_{WL} and P_{NW}. It combines the advantage of Narrow Lane LC and Wide Lane LC resulting useful to make easier the detection of cycle sleeps:

$$L_{MW} = \frac{f_1 L1 - f_2 L_2}{f_1 - f_2} - \frac{f_1 P1 + f_2 P_2}{f_1 + f_2} \tag{2.27}$$

2.3.13.2 Differential observables

Considering two receivers (k and m) and one satellite (p), single difference pseudorange and phase observables can be generated by difference with respect to the two receivers[4]:

$$\begin{aligned} \rho^p_{km}(t_k, f) &= \rho^p_k(t_k, f) - \rho^p_m(t_k, f) \\ \delta\phi^p_{km}(t_k, f) &= \delta\phi^p_k(t_k, f) - \delta\phi^p_m(t_k, f) \end{aligned} \tag{2.28}$$

where the pseudorange and carrier phase notations are the same of Eqs. 2.6 and 2.7. The two receivers measurements are supposed fully synchronised (referred to the same time $t_k = t_m$), and the time difference between the satellite transmission instants are negligible ($t^p_k \simeq t^p_m$). The satellite-based error sources cancel out by making single difference with respect to a single satellite. The remaining terms are geometric, atmosphere dependent, and based on receiver error sources:

$$\begin{aligned} \rho^p_{km}(t_k, f) &= R^p_{km}(t_k) + c[\epsilon_{r\ km}(t_k) + \Delta T^p_{Trop\ km}(t_k) + \Delta T^p_{Ion\ km}(t_k, f) + \\ &+ \Delta T_{Ric\ km}(t_k, f) + \Delta T^p_{Multi\ km}(t_k, f)] \\ \delta\phi^p_{km}(t_k, f) &= \frac{f}{c} R^p_{km}(t_k) + f[\epsilon_{r\ km}(t_k) + \Delta T^p_{Trop\ km}(t_k) + \Delta T^p_{Ion\ km}(t_k, f) + \\ &+ \Delta T_{Ric\ km}(t_k, f) + \Delta T^p_{Multi\ km}(t_k, f)] - N^p_{km}(t_k, f) \end{aligned} \tag{2.29}$$

Considering two receiver (k and m) and two satellites (p and q), double difference pseudorange and phase observables can be generated by difference of single differences (as shown in Eq. 2.28) with respect to two satellites [5]:

$$\begin{aligned} \rho^{pq}_{km}(t_k, f) &= \rho^p_{km}(t_k, f) - \rho^q_{km}(t_k, f) \\ \delta\phi^{pq}_{km}(t_k, f) &= \delta\phi^p_{km}(t_k, f) - \delta\phi^q_{km}(t_k, f) \end{aligned} \tag{2.30}$$

In this combination also the receiver-based errors cancel out, and the other errors are greatly reduced:

$$\begin{aligned} \rho^{pq}_{km}(t_k, f) &= R^{pq}_{km}(t_k) + c[\Delta T^{pq}_{Trop\ km}(t_k) + \Delta T^{pq}_{Ion\ km}(t_k, f) + \Delta T^{pq}_{Multi\ km}(t_k, f)] \\ \delta\phi^{pq}_{km}(t_k, f) &= \frac{f}{c} R^{pq}_{km}(t_k) + f[\Delta T^{pq}_{Trop\ km}(t_k) + \Delta T^{pq}_{Ion\ km}(t_k, f) + \\ &+ \Delta T^{pq}_{Multi\ km}(t_k, f)] - N^{pq}_{km}(t_k, f) \end{aligned} \tag{2.31}$$

[4] Considering the generic quantity V^p dependent on the position of the satellite p we introduce the notation V^p_{km} to indicate the difference of the values measured by the k and m receivers: $V^p_{km} = V^p_k - V^p_k$.

[5] Being V^p_{km} the single difference between receivers k and m of the quantity V^p, the double difference is simply given by: $V^{pq}_{km} = V^p_{km} - V^q_{km}$.

By eliminating the satellite and receiver-based error sources the precision of the Differential GPS (DGPS) positioning improves. As might be expected the precision of double differences observables is greater with respect to single difference.

To reach very high precision of measurement the difference of two double differences over time (t_k and $t_k + t_0$) can be made. The triple difference observable benefits of eliminating the integer ambiguity of carrier phase ($N^{pq}_{km}(t_k, f)$ constant along the time). Moreover if the (t_0, $t_0 + t_k$) time interval is small most of the quantities of Eq. 2.31 cancel out.

$$\begin{aligned}
\Delta\rho^{pq}_{km}(t_k, t_k + t_o, f) &= \rho^{pq}_{km}(t_k, f) - \rho^{pq}_{km}(t_k + t_o, f) = \\
&\simeq R^{pq}_{km}(t_k) - R^{pq}_{km}(t_k + t_0) \\
\Delta\delta\phi^{pq}_{km}(t_k, t_k + t_o, f) &= \delta\phi^{pq}_{km}(t_k, f) - \delta\phi^{pq}_{km}(t_k + t_o, f) = \\
&\simeq \frac{f}{c}[R^{pq}_{km}(t_k) - R^{pq}_{km}(t_k + t_0)]
\end{aligned} \tag{2.32}$$

Triple difference are very useful for cycle sleep detection, as they occur as sudden spike in the time trend of values.

By making the differences of measurements of one receiver relative to two different satellites (p and q) it can be generated the observable known as between satellite single difference. In this observable the receiver-based errors cancel out.

2.4 The GLONASS and the Galileo systems

2.4.1 GLONASS

The Russian Global'naya Navigatsionnaya Sputnikkovaya Sistema (GLONASS), developed since the launch of the first satellite of its constellation in 1982, is an alternative to the US GPS system. As GPS system it is composed by two main services: the military one and the civil one, with civil services (and application quality) degraded with respect to military ones. The full constellation is composed by 24 satellites (excluding spare ones) equally spaced over 3 orbital planes. Satellite orbits are near circular at an altitude of 19100 km over the Earth surface. Orbital planes are inclined of 64.8 degrees with respect to the equatorial plane. Revolution time is approximately 11 hours and 15 minutes. The ground control stations of the GLONASS are maintained only in the territory of the former Soviet Union, thus limiting the global coverage capability of the system. The control segment provides the monitoring of the full constellation, and the upload of orbital parameters for broadcasting. The satellite coverage of GLONASS is greater at northern latitudes, where the GPS system has its minimum coverage. In the signal transmission the GLONASS system uses a spread spectrum technology; the signal are right-hand circularly polarised and the modulation type is the BPSK. The code is the same for all satellites, but each satellite transmit in a different frequency

using a total of 25-channels to realise a Frequency Division Multiple Access (FDMA)[6].

Two main frequencies are used for navigation: L1 (1602.0 MHz) and L2 (1246 MHz). The central frequency relative to each satellite carrier can be obtained as follow:

$$\begin{aligned} L1 &= 1602 + n \times 0.5625 \text{ MHz} \\ L2 &= 1246 + n \times 0.4375 \text{ MHz} \end{aligned} \tag{2.33}$$

where n is the frequency channel. To avoid interference problems using frequencies with mobile satellite terminals, the channels have been set to $n = -7, ..., 0, ..., 6$. The L1 and L2 carrier frequencies are coherently obtained from a time/frequency reference standard on board of each satellite. The nominal value of the reference frequency is 5.0 MHz at ground[7]

GLONASS satellites transmit two types of signals: a Standard Precision (SP) signal and a so called "obfuscated" High Precision (HP) signal. The chipping rates for the HP code and the SP code are 5.11 and 0.511 Mbps, respectively. The HP signal is broadcasted in quadrature phase with the SP signal, it has a ten time larger bandwidth and it is available only for authorised users. As well as for GPS system also in GLONASS it is possible to make phase measurements by looking the carrier signal. The navigation message (50 bps) contains the parameters to compute spatial and temporal position and speed of each satellite. The updating frequency of satellites ephemeris is 30 minutes. The combination of measurements and navigation messages allows to determine the position coordinates, the speed vectors and the time of the receiver.

The modulating signals for L1 and L2 carrier frequency are generated as a modulo-2 addition of binary signal sequences of the SP ranging code transmitted at 511 kbps the navigation message at 50 bps (only in the L1 frequency) and a 100 Hz auxiliary sequence.

GLONASS has been operational since 1982; the first modernisation plan is the GLONASS-M since 2003, that in 2011 has allowed to reach the full constellation availability. The second modernisation plan is being implemented: GLONASS-K from 2011, and at 06:07 Moscow Time on 26 February 2011, the first GLONASS-K satellite was launched. Civilian GLONASS used alone is very slightly less accurate than GPS. The GLONASS-K satellite however will double the accuracy of the previous system. The GLONASS-K satellites constellation will introduce the CDMA signal for testing purpose, and a new carrier frequency. This will allow to create a frequency overlapping with the GPS system and the future Galileo system as well as to perform a better ionosphere effect correction using three frequencies.

[6]Satellites placed to the antipodes of an orbital plane, that are never visible to the same user, are set to use the same frequency, in order to avoid interference with mobile satellite terminals.

[7]In order to compensate for relativistic effects, the satellite nominal frequency value is set to 4.99999999782 MHz, analogously to what is done in the GPS system (see § 2.3.5).

GPS and GLONASS integration must overcome the difficulties of using different spatial reference systems (ECEF for GPS and PZ-90 for GLONASS), and different time reference system. The transformation between different systems introduces errors in the accuracy of measurements. Furthermore a GLONASS GPS receiver must be equipped with two different hardware, rising its cost. However, once overcome the technical difficulties, very high accuracy could be achieved, also considering the continuous systems evolution.

2.4.2 Galileo

GALILEO is a joint initiative of European Space Agency (ESA) and European Union (EU). It is a global satellite navigation system designed to provide a multimodal service, in different domains. It is conceived to be completely independent and autonomous, but consistent and interoperable with the american GPS system and the russian GLONASS. Galileo system is mainly a civil initiative and will develop different level civil applications:

- Open Service (OS)
- Commercial Service (CS)
- Safety Of Life (SOL)
- Public Regulated Service (PRS)
- Search & Rescue (SAR)

The Space segment consists of a constellation of 30 MEO satellites (27 operational + 3 active spares) with orbit at height of 23616 km. Satellites are planned to be placed in 3 orbital planes with an inclination of 56 degrees over the terrestrial equatorial plane. The orbital period is 14 hours and 22 minutes.
The Galileo system is designed to send signals in four spectral bands:

- E5a: $1164 - 1191.795$ MHz
- E5b: $1191.795 - 1215$ MHz
- E6: $1260 - 1300$ MHz
- E1: $1559 - 1591$ MHz

The available carrier signals are:

- E1: 1575.420 MHz
- E6: 1278.750 MHz

- E5: 1191.795 MHz
- E5a: 1176.450 MHz
- E5b: 1207.140 MHz

Similarly to GPS, the Galileo system will be provided of Ranging Codes (RC). To allow the completion of all services each Galileo satellite will broadcast 10 different navigation signals and 1 Search And Rescue signal. Six signals are accessible to all Galileo Users on the E5a, E5b and L1 carrier frequencies for OS and SoL services: E5a (I); E5a (Q); E5b (I); E5b (Q); E2-L1-E1 (B); E2-L1-E1(C). Two signals on E6 with encrypted ranging codes are accessible only for CS provider: E6 (B); E6 (C). Finally, two signals are accessible to authorised users of PRS: E6 (A); E2-L1-E1(A). Different

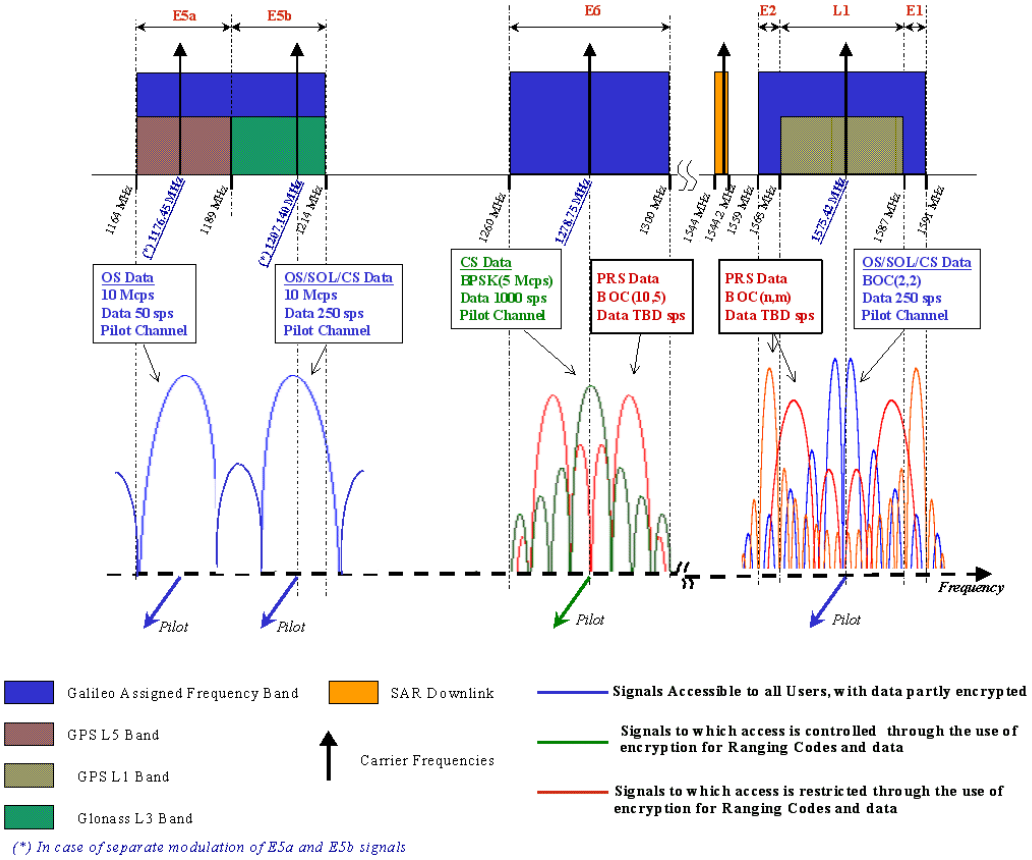


Figure 2.8: Structure of the Galileo Navigation Signal ([26]).

codes for different satellite reception is possible using CDMA. Different signals are broadcasted on the in-phase (I) and in quadrature (Q) channels and, in the case of the E5, different signals are provided in the upper (E5b) and lower (E5a) part of the band.

The first experimental satellites, GIOVE-A and GIOVE-B have been launched in 2005 and in 2008 respectively. Their mission was the test of satellite environment and instrumentation. Test on the signal-in-space have been made on the frequency spectrum required for Galileo.

Following four IOV (In Orbit Validation) Galileo satellites have been launched two in october 2011 and two october 2012, giving in march 2013 the very first position determination of a ground location. These satellites are much closer to the final satellite system configuration. The initial satellite navigation services will be available in 2014. A series of 14 further satellite will allow to reach the intermediate operational capability configuration in 2015 and finally the remaining 12 satellites will be deployed to reach the Full Operational Capability (FOC) by 2020.

2.4.3 Complementarity and interoperability

As shown in Fig. 2.8 the Galileo system uses frequency bands that are partially overlapped to the future GLONASS L3 and the GPS L5 bands. Moreover future GLONASS satellite will provide CDMA signals as the GPS does. This is an indication of a design of an interoperable GNSS, developed for market requirements. The primary requirement is to allow interoperability at the system level. A GNSS receiver hardware must be designed to enable the simultaneous tracking and decoding of the different constellations signals, ensuring the nominal accuracy of the systems themselves. On the other hand the design of new generations of satellite navigation signals, is such to ensure the possibility of a simplification of the receiver hardware and software to make it able to easily decode and process all signals simultaneously. Last generations of mobile smartphones are already able to simultaneously receive and decode civilians signals coming from both GPS and GLONASS satellites, and will also easily incorporate Galileo as soon as it becomes operational.

The compatibility and interoperability of satellite navigation systems will involve the different reference frame and their definitions. To make interoperable the reference frame they need to be "very close". As an example the difference between the Galileo Terrestrial Reference Frame (GTRF) and the World Geodetic System 1984 (WGS84) ECEF is expected to be about 3 cm.

The time reference offered by satellite navigation is another aspect of the compatibility of the systems. A solution can be to provide the parameters for systems synchronisation, or to equip the receiver with a software for the synchronisation computation.

2.5 User position computation

In the GPS system the user position is referred to an Earth rotating reference frame, the ECEF cartesian coordinate system, to facilitate the computation of geographical coordinates (latitude, longitude, heigh) of the receiver position. The ECEF reference system has the origin in the Earth centre of mass, the x-y plane coincident with the equatorial plane, the x axis oriented along 0 degrees longitude, and the z axis normal

to the x-y plane oriented to the north pole.

The user position computation algorithm using broadcasted ephemeris is briefly described in this section.

2.5.1 Time Measurement

The most diffuse time reference frame is the UTC (Universal Time Coordinated) or Greenwich Mean Time (GMT) which uses atomic clocks measurements to take synchronisation and is based on the mean solar day duration. The Earth rotation around its spin axis slows down due to some factors: astronomical (joint moon sun gravitational pull), in-Earth (i.e Earth's kernel and mantle movements and perturbations), out-Earth (i.e atmosphere and ocean tides). This effect of Earth rotation slowing increases the duration of the mean solar day (from 86400 s to 86400, 002 s). During a year this difference accumulates. To have the sun over the Greenwich meridian at noon, with very good approximation, this difference is maintained within 0.9 s by adding, when necessary, 1 s at midnight at the end of the year. This quantity is known as leap seconds. The International Earth Rotation and Reference System Service (IERS) has the task of determining the updating of the leap seconds quantity.

The International Atomic Time (TAI) is a time scale very accurate and stable, available since 1955, which uses a statistical approach in the time measurement by involving an extensive number of atomic clocks, positioned around the world and maintained by the "Bureau International des Poids et Mesures". The TAI time is not adjusted by the leap seconds quantity, to maintain atomic clock synchronisation.

In the GPS system the time measurement is performed using a proper time reference frame. The GPS epoch is 0000 UTC (midnight) on January 6, 1980. Since this start time, the time is measured in weeks and seconds since the start of the week (midnight between saturday and sunday). The GPS time doesn't take into account for leap seconds, as the satellites operates in a continuous time scale. Its is important to note that leap seconds aligning TAI and GPS time measurements are not the same aligning TAI and UTC. In particular at time of the present work (2013) there are the following relationships between different time reference systems (TAI, UTC, GPS):

$$\begin{aligned} TAI &= UTC + 35s \\ TAI &= GPS + 19s \\ GPS &= UTC + 16s \end{aligned} \tag{2.34}$$

The synchronisation is maintained by the MCS using Earth-based atomic clocks around the globe and on board of satellites. At the MCS the computation of each satellite clock bias is modeled and uploaded to the satellite, for broadcasting in the navigation message (§ 2.5.2).

2.5.2 The Ephemeris

Each satellite transmit the orbital parameters in the navigation message. The parameters are referred to a determined time (epoch); below they are listed following the ICD ([21]):

t_{oc}: clock reference time (in seconds) used for clock offset computation;

a_0 , a_1, a_2: polynomial coefficient for clock offset computation;

t_{GD}: (Group Delay), instrumental differential delay;

t_{oe}: reference time (in seconds) since the GPS week start (at the Saturday/Sunday transition) of ephemeris values;

M_0: mean anomaly;

Δn: mean motion difference from computed value;

e: satellite orbit eccentricity;

$\sqrt{a}$: square root of orbital semi-major axis;

Ω_0: latitude of ascending node of the orbit plane at the weekly epoch;

i_0: inclination angle of the orbit plane with respect to the equatorial plane;

ω: argument of perigee;

$\dot{\Omega}_0$ rate of right ascension;

$\dot{i}$ rate of inclination angle;

C_{rc}: amplitude of the cosine harmonic correction term to the orbit radius;

C_{rs}: amplitude of the sine harmonic correction term to the orbit radius;

C_{uc}: amplitude of the cosine harmonic correction term to the argument of latitude;

C_{us}: amplitude of the sine harmonic correction term to the argument of latitude;

C_{ic}: amplitude of the cosine harmonic correction term to the angle of inclination;

C_{is}: amplitude of the sine harmonic correction term to the angle of inclination.

2.5.3 Correction of the transmission time

Starting from the time of signal reception (t_j) the uncorrected transmission time (t^i_{ju}) from the i-th satellite is given by[8]:

$$t^i_{ju} = t_j - \frac{\rho^i_j}{c} \tag{2.35}$$

where ρ^i_j is the pseudorange measurement of the j-th receiver at time t_j and c is the speed of light.

The transmission epoch must be corrected for the satellite clock offset and for the periodic relativistic effect. This can be done using ephemeris data. The satellite mean motion (n) is given by:

$$n = \sqrt{\frac{\mu}{a^3}} + \Delta n \tag{2.36}$$

where $\mu = GM = 3.9860005 \times 10^{14}$ m^3 s^{-2} is the Earth gravitational constant (in the WGS84 system). The mean M anomaly can be corrected using the value of n:

$$M = M_0 + n(t^i_{ju} - t_{oe}) \tag{2.37}$$

The eccentric anomaly E must be computed by solving the Kepler equation, for example using an iteration method (see § 2.5.3.1):

$$E = M + e \sin E \tag{2.38}$$

Introducing the constant $F = \frac{-2\sqrt{\mu}}{c^2} = -4.442807633 \times 10^{-10}$, the relativistic correction term is:

$$\Delta t_r = Fe\sqrt{a} \sin E \tag{2.39}$$

The parameters to correct the satellite clock offset ($a_{f0}, a_{f1}, a_{f2}, T_{GD}$) are broadcasted by the satellite in the navigation signal. The final correction formula is:

$$\Delta t = a_{f0} + a_{f1}(t_c - t_{oc}) + a_{f2}(t_c - t_{oc})^2 + \Delta t_r - T_{GD} \tag{2.40}$$

and the transmission epoch can be corrected:

$$t^i_j = t^i_{ju} - \Delta t \tag{2.41}$$

2.5.3.1 Solution of Kepler equation

In this section we describe the adopted solution to solve the Kepler equation:

$$E = M + e \sin E \tag{2.42}$$

We implemented an iterative algorithm to solve the equation:

$$E_k = M + e \sin E_{k-1} \tag{2.43}$$

[8]The same notation as § 2.3.1 and § 2.3.1 is used, superscripts indicate the satellite and subscripts indicate the receiver. The subscript $_u$ indicates the an uncorrected value.

Starting from $E_0 = M$ the successive steps are:

$$\begin{aligned} E_1 &= M + e\sin E_0 = M + e\sin M \\ E_2 &= M + e\sin E_1 = M + e\sin\left(M + e\sin M\right) \\ E_3 &= M + e\sin E_2 = M + e\sin\left[M + e\sin\left(M + e\sin M\right)\right] \\ E_4 &= M + e\sin E_3 = M + e\sin\left\{M + e\sin\left[M + e\sin\left(M + e\sin M\right)\right]\right\} \end{aligned} \quad (2.44)$$

Defining the function: $f(x) = x - e\sin x - M$ where x is an unknown quantity, the solution of $f(x) = 0$ is $x = E$. Defining ϵ as the error made by using the value of x as an approximation of E ($\epsilon = x - E$), and applying the Taylor expansion around the value $E = x - \epsilon$, and assuming ϵ small, we have:

$$f(x) = \sum_{i=0}^{+\infty} \frac{f^{(n)}(E)}{n!}(x - E)^n \Longrightarrow f(x - \epsilon) = \sum_{i=0}^{+\infty} \frac{f^{(n)}(x)}{n!}(-\epsilon)^n \quad (2.45)$$

around the value E. Remembering that:

$$\begin{aligned} f'(x) &= 1 - e\cos x; \\ f''(x) &= e\sin x; \\ f'''(x) &= e\cos x; \end{aligned} \quad (2.46)$$

Developing the analytical steps:

$$f(E) = f(x - \epsilon) = x - e\sin x - M - (1 - e\cos x)\epsilon + \frac{1}{2}\epsilon^2 e\sin x - \frac{1}{6}\epsilon^3 e\cos x \quad (2.47)$$

Limiting the expansion to first order, and extracting the value of ϵ solution of the equation $f(x - \epsilon) = 0$:

$$\epsilon = \frac{x - e\sin x - M}{1 - e\cos x} \quad (2.48)$$

If we consider the expansion limiting to the second order:

$$\epsilon = \frac{x - e\sin x - M}{1 - e\cos x - \frac{1}{2}\epsilon e\sin x} \quad (2.49)$$

Using the equation (2.48) to substitute ϵ in the denominator of (2.49):

$$\epsilon = \frac{x - e\sin x - M}{1 - e\cos x - \frac{1}{2}e\sin x\frac{x - e\sin x - M}{1 - e\cos x}} \quad (2.50)$$

Finally in a similar way, using also the third order of the expansion:

$$\epsilon = \frac{x - e\sin x - M}{1 - e\cos x - \frac{1}{2}(e\sin x - \frac{1}{3}e\cos x\epsilon)\epsilon} \quad (2.51)$$

We solve the Kepler equation by iteration, starting from the first guess: $x = x_0$. The value used for the initialisation of the algorithm is (see above):

$$x_0 = M + e\sin\left\{M + e\sin\left[M + e\sin\left(M + e\sin M\right)\right]\right\} \quad (2.52)$$

The following steps are subsequently iterated:

$$\epsilon_n = \frac{x_n - e\sin x_n - M}{1 - e\cos x_n - \frac{1}{2}e\sin x_n \frac{x_n - e\sin x_n - M}{1 - e\cos x_n}} \tag{2.53}$$

Then the value of ϵ_n found by (2.53) can be used to compute the value of:

$$\epsilon_{n+1} = \frac{x_n - e\sin x_n - M}{1 - e\cos x_n - \frac{1}{2}(e\sin x_n - \frac{1}{3}e\cos x_n\epsilon_n)\epsilon_n} \tag{2.54}$$

So the estimated value of E at step n+1 is $E_{n+1} = x_n - \epsilon_{n+1}$. Then we use this value E_{n+1} as a starting value for a new step of the algorithm until the needed precision ($dE = |E_{n+1} - x_n|$) is not reached.

To avoid an infinitive loop we have fixed a maximum step number of 100. If the loop is repeated 100 times the procedure is automatically interrupted.

2.5.4 The satellite position

The satellite position is univocally determined starting from the time instant (epoch) and the orbital parameters (ephemeris), broadcasted by the satellite in the navigation message. Ephemeris are updated by MCS every 2 hours. The precision of the satellite position, using broadcasted ephemeris is of the order of meters. The precision of satellite position can be improved using precise ephemeris (see § 2.3.7 for more details). The true anomaly (ν) is computed using a formula that accounts for the quadrant ambiguity by using the tangent function:

$$\tan\frac{\nu}{2} = \sqrt{\frac{1+e}{1-e}}\tan\frac{E}{2} \tag{2.55}$$

Argument of latitude (ϕ), radius(r), inclination(i), and longitude of ascending node (Ω) must be corrected for the effects of not perfect ellipticity, using correction terms directly sent in the navigation signal (see § 2.5.2):

$$\begin{aligned}
\phi = \phi + \delta(\phi) \Longrightarrow \quad \phi &= \omega + \nu + C_{uc}\cos 2(\omega+\nu) + C_{us}\sin 2(\omega+\nu) \\
r = r + \delta(r) \Longrightarrow \quad r &= a(1 - e\cos E) + C_{rc}\cos 2(\omega+\nu) + C_{rs}\sin 2(\omega+\nu) \\
i = i + \delta(i) \Longrightarrow \quad i &= i_0 + \dot{i}(t - t_{oe}) + C_{ic}\cos 2(\omega+\nu) + C_{is}\sin 2(\omega+\nu) \\
\Omega &= \Omega_0 + (\dot{\Omega}_0)(t - t_{oe}) - \dot{\Omega}_e t
\end{aligned} \tag{2.56}$$

where $\dot{\Omega}_e = 7.292115167 \times 10^{-5}$ rad s^{-1} is the Earth rotation speed (in the WGS84 system).

The position of the satellite is in a coordinate system ($GX_EY_EZ_E$) where the origin is the Earth centre of mass (G), the X_EY_E plane is the orbital plane, with X_E passing through the ascending node and Y_E perpendicular to X_E; finally Z_E is perpendicular to the X_EY_E plane. The obtained tern must be clockwise. Transformations to obtain

(X_E, Y_E, Z_E) are:

$$\begin{aligned} X_E &= r\cos(\phi) \\ Y_E &= r\sin(\phi) \\ Z_E &= 0 \end{aligned} \tag{2.57}$$

Finally to express the position of the satellite in the ECEF reference system:

$$\begin{aligned} X &= X_E\cos(\Omega) - Y_E\cos(i)\sin(\Omega) \\ Y &= X_E\sin(\Omega) + Y_E\cos(i)\sin(\Omega) \\ Z &= Y_E\sin(i) \end{aligned} \tag{2.58}$$

In order to compute the user position, the receiver and satellite position must be referred to the same reference frame. As shown in § 2.3.8 some transformations must be introduced to take into account for the Earth rotation effects during signal travel time (Sagnac effect): these transformation expressed in Eq. 2.21 can be directly applied to the satellite position coordinates.

2.5.5 Position computation: least square estimation

The pseudorange measurement is affected by a series of errors (Eq. 2.6) that must be corrected as accurately as possible to obtain the precise location of the receiver (§ 2.3.1). Some correction terms ($\epsilon_s^i(t_j^i)$, $\Delta T_{Sat}^i(t_j^i)$, $\Delta T_{Orbit}^i(t_j^i)$, $\Delta T_{Rel}^i(t_j^i)$) can be directly computed by applying algorithm on parameters broadcasted in the navigation message. Some other terms ($\Delta T_{Ric}^i(t_j)$, $\Delta T_{Ion}^i(t_j, f)$, $\Delta T_{Trop}^i(t_j)$, $\Delta T_{Multi(t_j,f)}^i$) must be directly corrected using models, and other strategies that we have previously described. Once all parameters are computed, the remaining terms are the geometric distance and the receiver clock offset which constitutes the fourth unknown together with the receiver coordinates in the problem of positioning. By expressing the geometric distance of the j-th receiver from the i-th satellite as:

$$R_j^i(t_j) = \sqrt{[X^i(t_j^i) - X_j(t_j)]^2 + [Y^i(t_j^i) - Y_j(t_j)]^2 + [Z^i(t_j^i) - Z_j(t_j)]^2} \tag{2.59}$$

and considering four satellite simultaneously in view, we can write a non-linear system of four equations in four unknowns, starting from the four corrected pseudoranges:

$$\begin{cases} \rho_j^1(t_j, f) = \sqrt{[X^1(t_j^1) - X_j(t_j)]^2 + [Y^1(t_j^1) - Y_j(t_j)]^2 + [Z^1(t_j^1) - Z_j(t_j)]^2} + c\epsilon_{rj}(t_j) \\ \rho_j^2(t_j, f) = \sqrt{[X^2(t_j^2) - X_j(t_j)]^2 + [Y^2(t_j^2) - Y_j(t_j)]^2 + [Z^2(t_j^2) - Z_j(t_j)]^2} + c\epsilon_{rj}(t_j) \\ \rho_j^3(t_j, f) = \sqrt{[X^3(t_j^3) - X_j(t_j)]^2 + [Y^3(t_j^3) - Y_j(t_j)]^2 + [Z^3(t_j^3) - Z_j(t_j)]^2} + c\epsilon_{rj}(t_j) \\ \rho_j^4(t_j, f) = \sqrt{[X^4(t_j^4) - X_j(t_j)]^2 + [Y^4(t_j^4) - Y_j(t_j)]^2 + [Z^4(t_j^4) - Z_j(t_j)]^2} + c\epsilon_{rj}(t_j) \end{cases} \tag{2.60}$$

In these equations the non linear term:

$$f(X_j, Y_j, Z_j) = \sqrt{[X^i(t_j^i) - X_j(t_j)]^2 + [Y^i(t_j^i) - Y_j(t_j)]^2 + [Z^i(t_j^i) - Z_j(t_j)]^2} \quad (2.61)$$

can be linearized starting from an initial receiver position value $(X_{j,0}, Y_{j,0}, Z_{j,0})$, sometimes coincident with the origin of the reference system. Then considering the increments $(\Delta X_j, \Delta Y_j, \Delta Z_j)$:

$$\begin{aligned} X_{j,1} &= X_{j,0} + \Delta X_j \\ Y_{j,1} &= Y_{j,0} + \Delta Y_j \\ Z_{j,1} &= Z_{j,0} + \Delta Z_j \end{aligned} \quad (2.62)$$

by Taylor expansion around the initial position and limiting to first order

$$\begin{aligned} f(X_{j,1}, Y_{j,1}, Z_{j,1}) &= f(X_{j,0}, Y_{j,0}, Z_{j,0}) + \frac{\delta f(X_{j,0}, Y_{j,0}, Z_{j,0})}{\delta X_{j,0}} \Delta X_j + \\ &+ \frac{\delta f(X_{j,0}, Y_{j,0}, Z_{j,0})}{\delta Y_{j,0}} \Delta Y_j + \frac{\delta f(X_{j,0}, Y_{j,0}, Z_{j,0})}{\delta Z_{j,0}} \Delta Z_j \end{aligned} \quad (2.63)$$

computing the partial derivatives:

$$\begin{aligned} \frac{\delta f(X_{j,0}, Y_{j,0}, Z_{j,0})}{\delta X_{j,0}} &= -\frac{X^i - X_{j,0}}{R_{j,0}^i} \\ \frac{\delta f(X_{j,0}, Y_{j,0}, Z_{j,0})}{\delta Y_{j,0}} &= -\frac{Y^i - Y_{j,0}}{R_{j,0}^i} \\ \frac{\delta f(X_{j,0}, Y_{j,0}, Z_{j,0})}{\delta Z_{j,0}} &= -\frac{Z^i - Z_{j,0}}{R_{j,0}^i} \end{aligned} \quad (2.64)$$

$$\begin{aligned} \rho_j^i &= R_{j,0}^i - \frac{X^i - X_{j,0}}{R_{j,0}^i} \Delta X_j - \frac{Y^i - Y_{j,0}}{R_{j,0}^i} \Delta Y_j \\ &- \frac{Z^i - Z_{j,0}}{R_{j,0}^i} \Delta Z_j + c[\epsilon_{rj}(t_j)] \end{aligned} \quad (2.65)$$

or

$$\begin{aligned} \rho_j^i - R_{j,0}^i &= -\frac{X^i - X_{j,0}}{R_{j,0}^i} \Delta X_j - \frac{Y^i - Y_{j,0}}{R_{j,0}^i} \Delta Y_j + \\ &- \frac{Z^i - Z_{j,0}}{R_{j,0}^i} \Delta Z_j + c[\epsilon_{rj}] \end{aligned} \quad (2.66)$$

introducing:

$$b_j^i = \rho_j^i - R_{j,0}^i \quad (2.67)$$

and considering the observables for all satellites simultaneously in view (n) a system of n equation in four unknowns with $n > 4$:

$$\mathbf{AX} = \mathbf{b} \quad (2.68)$$

$$
\begin{aligned}
\mathbf{A} &= \begin{pmatrix} -\frac{X^1-X_{j,0}}{R^1_{j,0}} & -\frac{Y^1-Y_{j,0}}{R^1_{j,0}} & -\frac{Z^1-Z_{j,0}}{R^1_{j,0}} & 1 \\ -\frac{X^2-X_{j,0}}{R^2_{j,0}} & -\frac{Y^2-Y_{j,0}}{R^2_{j,0}} & -\frac{Z^2-Z_{j,0}}{R^2_{j,0}} & 1 \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{X^n-X_{j,0}}{R^n_{j,0}} & -\frac{Y^n-Y_{j,0}}{R^n_{j,0}} & -\frac{Z^n-Z_{j,0}}{R^n_{j,0}} & 1 \end{pmatrix} \\
\mathbf{X} &= \begin{pmatrix} \Delta X_j \\ \Delta Y_j \\ \Delta Z_j \\ c[\epsilon_{rj}] \end{pmatrix} \\
\mathbf{b} &= \begin{pmatrix} b^1_j \\ b^2_j \\ \vdots \\ b^n_j \end{pmatrix}
\end{aligned} \tag{2.69}
$$

Being the number of observations (n) greater than the number of unknowns (4), the solution of the system of equations can be made by finding an estimation of the $\mathbf{X}$ vector, denoted as $\hat{\mathbf{X}}$. The error vector: $\hat{\mathbf{e}} = \mathbf{b} - \mathbf{A}\hat{\mathbf{X}}$ can be minimised using the least square estimation method:

$$
\begin{aligned}
\|\hat{\mathbf{e}}\|^2 &= (\mathbf{b} - \mathbf{A}\hat{\mathbf{X}})^T(\mathbf{b} - \mathbf{A}\hat{\mathbf{X}}) = (\mathbf{b}^T - \mathbf{A}^T\hat{\mathbf{X}}^T)(\mathbf{b} - \mathbf{A}\hat{\mathbf{X}}) = \\
&= \mathbf{b}^T\mathbf{b} - \mathbf{b}^T\mathbf{A}\hat{\mathbf{X}} - \mathbf{A}^T\mathbf{b}\hat{\mathbf{X}}^T - \mathbf{A}^T\mathbf{A}\hat{\mathbf{X}}^T\hat{\mathbf{X}} = \\
&= \|\mathbf{b}\|^2 + \mathbf{A}^T\mathbf{A}\|\hat{\mathbf{X}}\|^2 - \mathbf{b}^T\mathbf{A}\hat{\mathbf{X}} - \mathbf{A}^T\mathbf{b}\hat{\mathbf{X}}^T
\end{aligned} \tag{2.70}
$$

To minimise $\|\hat{\mathbf{e}}\|^2$ we set to zero the derivative of the previous:

$$
\frac{\delta\|\hat{\mathbf{e}}\|^2}{\delta\hat{\mathbf{x}}} = 2\mathbf{A}^T\mathbf{A}\hat{\mathbf{x}} - \mathbf{b}^T\mathbf{A} - \mathbf{A}^T\mathbf{b} = 2\mathbf{A}^T\mathbf{A}\hat{\mathbf{x}} - 2\mathbf{A}^T\mathbf{b} \tag{2.71}
$$

being $\mathbf{b}^T\mathbf{A} = \mathbf{A}^T\mathbf{b}$. The derivative is zero if

$$
\mathbf{A}^T\mathbf{A}\hat{\mathbf{x}} = \mathbf{A}^T\mathbf{b} \Rightarrow \hat{\mathbf{x}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b} \tag{2.72}
$$

the corresponding value of $\hat{\mathbf{x}}$ is given by:

$$
\hat{\mathbf{x}} = \begin{pmatrix} \Delta X_{j,1} \\ \Delta Y_{j,1} \\ \Delta Z_{j,1} \\ c[\epsilon_{rj,1}] \end{pmatrix} \tag{2.73}
$$

the receiver position can be consequently updated:

$$
\begin{aligned}
X_{j,1} &= X_{j,0} + \Delta X_{j,1} \\
Y_{j,1} &= Y_{j,0} + \Delta Y_{j,1} \\
Z_{j,1} &= Z_{j,0} + \Delta Z_{j,1}
\end{aligned} \tag{2.74}
$$

Some iterations can be made by substituting these values to recompute the elements of the $\mathbf{A}$ array and the $\mathbf{b}$ vector. The new solution is obtained using the least square

estimation (2.72). The iteration can be made until the values of $\hat{\mathbf{x}}$ reach the order of magnitude of few meters (the nominal accuracy of GNSS pseudorange undifferenced positioning service).

2.5.5.1 Positioning precision

As an example some tests results of positioning are depicted in Fig. 2.9. They are obtained using the described algorithm applied only to pseudorange measurements. Corrections are performed with the parameters contained in the navigation message directly broadcasted by satellites. The corrected terms in the position determination have been: ionosphere effects, satellite clock offset, relativistic effects, DCBs and the Sagnac effect. The resulting error is due to the unmodeled errors such as residual orbit parameters errors (ephemeris used are those of the navigation message), multipath effect, antenna phase centre variation, ocean loading, solid Earth tides, polar motions and troposphere delay. A further error source, introducing a bias, is due to the fact that the position considered as true was taken directly from the header of the Receiver Independent Exchange Format (RINEX) files containing the pseudorange observations. In fact the accuracy of this position coordinate is not specified, and could be at meter levels. Generally the fixed receiver position is not taken from the RINEX header file, but is computed to achieve the needed precision. However, we only used it for a test of the implemented algorithm.

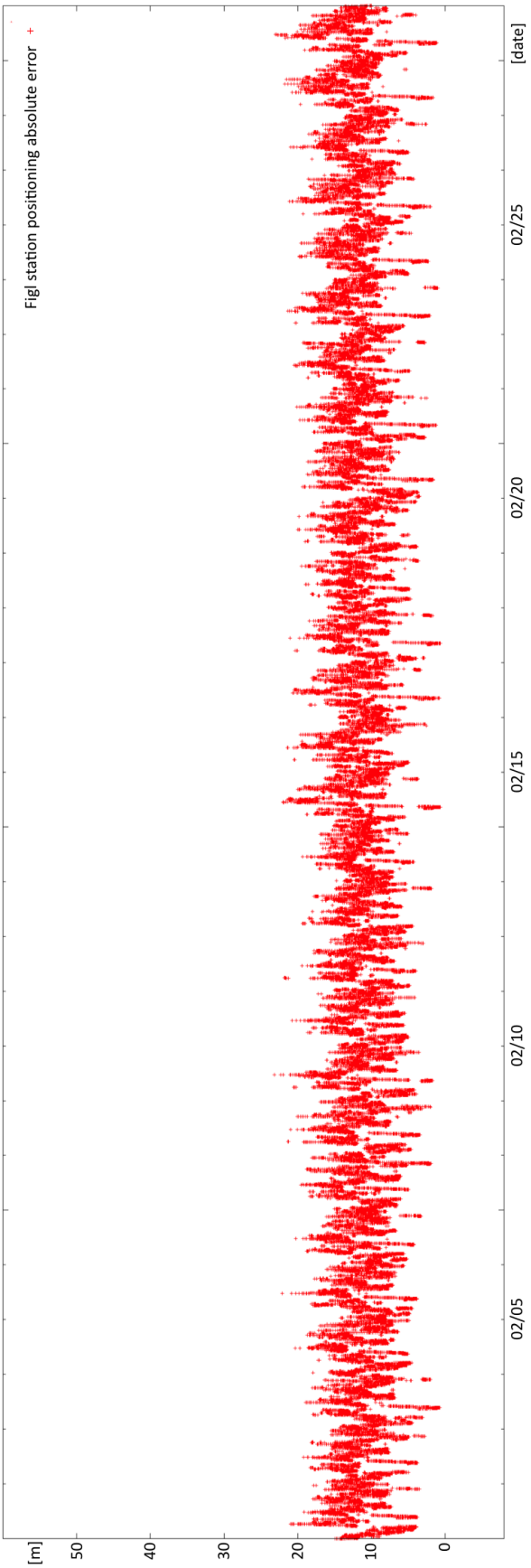


Figure 2.9: Time series of the absolute values for the geometric distances between the receiver position estimated by GPS positioning algorithm and the actual receiver position (taken in the RINEX files provided by the station). Station is figl (Figline Tuscany). Data are relative to the month of February in 2010.

Chapter 3

An overview of GNSS meteorology

The GNSS is used for a wide range of applications spanning from scientific to technological and commercial activities. In addition, the simplicity of use and low cost, encourage the installation of receivers at all latitudes, even in remote and isolated places such as the polar regions, resulting in the growth and spread of networks of GNSS stations. GPS Meteorology and more generally GNSS Meteorology (after the deployment of various constellations of the different satellite positioning systems), refers to the use of GNSS station data to extract the information of atmospheric parameters to allows to conduct studies of meteorological and climatological conditions, commonly also through the integration ground-based and satellite-based observations.

3.1 Techniques for the computation of the tropospheric delay

As we have seen in § 2 the positioning service of satellite navigation systems depends on some factors which must be taken into account, estimated and corrected with the required accuracy. For this reason the computation of all these factors, including the tropospheric influence on the GNSS signal was addressed from the beginning of the GNSS era. The first studies and attempts to estimate the tropospheric delay and then ZTD (or ZPD) are contemporary with the GPS system start of operations. Simultaneously the possibility of extract integrated values of water vapour content, given measurements of ground pressure spatially and temporally coincident with the ZPD measurement, has been investigated. The accuracy of water vapour content estimate is directly linked to the precision of the ZPD estimate.

Most of available geodetic softwares are able to process data coming from single receivers as well as from GNSS networks and they provide the estimate of ZTD with very high precision. At present most popular softwares are BERNESE ([27]), GAMIT ([28]) and GIPSY-OASIS ([29]).

As shown in § 2.3 the measurement of pseudorange (Eq. 2.6) and carrier phase (Eq. 2.7) includes the tropospheric effect on the signal. To extract the tropospheric

delay value all the other terms of the equation must be estimated. Some standard processing algorithm are made available by the GNSS community (see for instance [21]), whose most important were briefly introduced in § 2.3. Different processing approaches can be adopted: Zero Difference (ZD), Single Difference (SD), Double Difference (DD) and Precise Point Positioning (PPP). In all the algorithms the observations of a receiver relative to the various satellites in view are simultaneous, and therefore are affected by the same bias and drift of the receiver clock. This redundancy allows to estimate and remove the clock error of the receiver using basic algorithms (e.g. least squares) or more complex methods like optimal estimation algorithms (e.g. Kalman filter). Furthermore to estimate (or eliminate) the satellite clock errors and biases, it is necessary to process data from networks of stations. In most of cases the ZD, SD and DD approaches lead to the same result, whereby estimating or eliminating clock parameters makes no differences. A different approach is the PPP where the satellite clock errors are taken as an input data (see for instance [23]) and used in the signal processing. Only the receiver clock estimation must be performed and consequently the approach can be used also for a single receiver data processing.

By eliminating all other factors from pseudorange and carrier-phase the Slant Total Delay (STD) also known as Slant Path Delay (SPD) can be obtained. STD is referred to a single satellite receiver direction and it is highly noisy and varying in time for each pseudorange and carrier-phase observation. Consequently STDs can be computed by averaging over a time series of observations, to reduces the noise. The most common practice is the introduction of geometric function for the projection of the STDs to the zenith direction. All the STDs relative to different satellites in view are projected to the zenith using this geometric function. Redundancy is the way for the estimation of ZTD, directly from the observations using least square or other kind of estimation approaches.

The geodetic softwares parameterise the STD as the product of a mapping function (i.e. a function taking into account each observation geometry) times the ZTD using the following expression:

$$STD = m_h(\theta^i_{ej}(t_j))ZHD(t_j) + m_w(\theta^i_{ej}(t_j))ZWD(t_j) \tag{3.1}$$

where $m_h()$ and $m_w()$ are mapping functions for the hydrostatic and wet component respectively, $\theta^i_{ej}(t_j)$ is the elevation angle of the i-th satellite respect to the j-th receiver at time t_j. Also ZWD and ZHD are computed at time t_j.

As shown in § 1.3.2 the ZHD can be computed using surface measurements of pressure coincident in space and time with the GNSS measurement of the receiver.

Some mapping functions are based on truncated forms of continued fractions, the

first of which was introduced in [30]:

$$m(\theta_e) = \frac{1}{\sin\theta_e + \frac{a}{\sin\theta_e + \frac{b}{\sin\theta_e + \frac{c}{\sin\theta_e}}}} \tag{3.2}$$

where a,b,c can be constants or linear function empirically determined using refractivity profiles generally obtained using Very Long Baseline Interferometer (VLBI) and balloon data; θ_e is the elevation angle. Several mapping function of the same type (continued fraction expansion) have been developed, with increasing accuracy. The most widely used are the Niell Mapping Function (NMF) ([31]) and the Global Mapping Function (GMF) ([32]): both functions depend only on the station coordinates and the day of the year. The NMF is a fraction truncated at the third order, and zenith normalised at the unity; it is not dependent on a specific meteorological parameter, but is based on time fluctuations of atmospheric mass. The coefficients have been computed using profiles of U.S Standard Atmosphere Supplements, 1966 ([33]). Also the GMF is empirically based on the computation of coefficients, but using Numerical Weather Model (NWM) data collected from the European Centre for Medium Range Weather Forecasts (ECMWF).

Some other mapping functions are based on data from NWMs: the Isobaric Mapping Functions (IMF) are determined from the height of the 200 hPa pressure level for the hydrostatic part and from temperature and humidity for the wet part which can be easily extracted from NWM on a global grid ([34]); the Vienna Mapping Functions 1 (VMF1) are rigorously determined from the refractivity profiles downloaded with the best resolution at selected sites ([35]). All mapping functions have a range of applicability, typically spanning from the zenith direction (elevation=90 deg) to a cut-off angle (the lowest usable elevation) variable from case to case. On the other hand a cut-off angle must be introduced to exclude from the processing the observations affected by strong multipath effects.

In the model of Eq. 3.1 the atmosphere is assumed to be horizontally layered and azimuthal symmetric. Azimuthal asymmetries may introduce significant errors in geodetic measurements where high precision is required. To get higher order accuracy, the refractivity can be modeled in function of the horizontal direction resulting in a correction term owing to the azimuthal asymmetry. The improved model is given by an azimuthally asymmetric component added to the symmetric one.

$$\begin{aligned} STD &= m_h(\theta^i_{ej}(t_j))ZHD(t_j) + m_w(\theta^i_{ej}(t_j))ZWD(t_j) + \\ &+ m_\Delta(\theta^i_{ej}(t_j))\cot\theta^i_{ej}(t_j)\left[G_N\cos\theta^i_{aj}(t_j) + G_E\sin\theta^i_{aj}(t_j)\right] \end{aligned} \tag{3.3}$$

where $\theta^i_{aj}(t_j)$ is the azimuth angle measured eastward from north, $G = G_N + G_E$ is the gradient vector, namely the horizontal asymmetry, and $m_\Delta()$ is the gradient mapping function. The projection of the asymmetric delay in the horizontal plane is given by

the dot product between the azimuth direction vector[1] and the gradient vector, thus it is greatest when the azimuth vector points along the gradient vector, and it is zero when the azimuth vector is perpendicular to the gradient vector. By multiplying times the cotangent of the elevation angle, the component along the zenith direction is given; finally the mapping function allows to project along the satellite line of sight. This may allow to use the same mapping function for all the different components. Some further improved versions of the model gradient replace the term $m_\Delta(\theta^i_{ej}(t_j))\cot\theta^i_{ej}(t_j)$ with mapping functions of the azimuthal angle. As a result the values of ZWD, and gradient components are the unknown to be estimated, given a set of observations.

Considering both ZWDs and gradients as a function of time is however not very practical, because the redundancy reduces with the introduction of gradient parameters, and further it reduces having to estimate three different parameters for each time step. The gradient model is better suitable when both kinds of parameters (ZWDs and gradients) are considered global or in the context of a specific area where a network of receivers is available, which is able to increase the number of observations.

3.2 The IGS ZPD product and its accuracy

The ZPD estimation algorithm was initially developed to address the need of tropospheric correction in position determination, but from some years the scientists are using this product in the field of meteorology and climatology, focusing on the impact in the determination of the content of water vapour in the atmosphere and on the possibility to use the product as a source of information to be assimilated within the numerical weather models (§ 3.5). The ZPD product is organised as a correction term, as initially it was used in the positioning software processing. The distribution of the product is well-established, and has been active since 1997. The IGS has several Analysis Centers that process near real time data from the IGS reference receivers network to produce the ZPD products. Each centre uses different local ground networks, with different algorithms, softwares and parameters in the data processing. Born from a natural need for homogenisation, since 2003 at Jet Propulsion Laboratory (JPL) a new ZPD product has been implemented and made available. It uses the GIPSY-OASIS software ([36]) to process all available IGS sites data. Basically the software adopts a PPP method. Two steps can be distinguished in the processing. First, precise GNSS satellite positions and clock corrections are determined from a globally distributed network of receivers. The associated products are the satellite positions and clock errors, directly provided by the IGS. Several types of products are available, with different time resolutions, accuracies and delivery times. Some of them can be available up to

[1] We refer to the azimuth direction vector as the projection of the satellite tracking direction in the horizontal azimuthal plane.

every 5 minutes. To get higher sampling rates they must be time interpolated. The second step is the processing of single receiver or local network data to get the needed estimation. This processing uses a cut-off angle of 7 deg to eliminate observations on satellites too low on the horizon and the associated problems of multipath. An NMF ([31]) is used as mapping function both for the wet and the hydrostatic components. As first guess an hydrostatic delay value varying with height (typically 2.3 m at sea level) and a wet delay value typically fixed to 0.1 m are used. An unique file for each reference station and for each day is produced. The time sampling of the ZPD estimates is 5 minutes. This product has been used for the present work and it has been acquired as part of NASA's Earth Science Data Systems and archived and distributed by the Crustal Dynamics Data Information System (CDDIS). A subset of this dataset has been selected, i.e. a time series from 2001 up to 2011 of five minutes ZPDs, for some test sites over Mediterranean area (see § 4.2 for more details).

As the precision of the retrieved water vapour (§ 3.3) is strongly dependent on the precision of the ZPD estimation, we have made some comparisons on the selected dataset. Balloon data have been used as reference measurements. Between the IGS reference sites available for this work (see § 4.2) the Cagliari one (cagl) has been chosen because balloon data are closer to the GNSS station and the height limit of the observations is greater, thus it potentially gives a more reliable comparison.

Therefore, a comparison between ZPD, GNSS-based and estimated from balloon soundings respectively, has been applied to a dataset of about 10 years (2001 to 2010). The value of ZPD calculated by balloons is simply given by the integral of the refraction index $n(z)$ or the refractivity $N(z)$ along the vertical path between the surface level ($z = 0$) and the height of satellite (z_s) over the Earth surface:

$$ZPD = \int_0^{z_s} n(z) - 1 \; dz = \int_0^{z_s} N(z) \times 10^{-6} \; dz \tag{3.4}$$

As upper limit of balloon (z_b) data is certainly below z_s, this can be split in two terms

$$ZPD = \int_0^{z_b} N(z) \times 10^{-6} \; dz + \int_{z_b}^{z_s} N(z) \times 10^{-6} \; dz \tag{3.5}$$

Expliciting the refractivity as in Eq. 1.24, and assuming the absence of water vapour above the upper limit of the radiosonde, which is in the specific case almost always greater than 25 km, above which the wet refractivity term can be neglected:

$$ZPD = \int_0^{z_b} N(z) \times 10^{-6} \; dz + \int_{z_b}^{z_s} (\frac{P_d}{T} Z_d^{-1}) \times 10^{-6} \; dz \tag{3.6}$$

The term that refers to the part of atmosphere above the balloon observation is only an hydrostatic component. Similarly to what described in § 1.3 and focusing on Eq. 1.34, it can be expressed as a function of pressure and heigth of the upper balloon level. The term of zenith path delay, which is related to the portion of the troposphere sensed

by the balloon, has been approximated as a discrete piecewise linear integration of the refraction index:

$$ZPD = \sum_{0}^{K_b-1} \frac{n_i + n_{i+1}}{2} \cdot (z_{i+1} - z_i) + \\ + 22.276 \times 10^{-6} \frac{P_{K_b}}{9.784(1 - 0.002626\cos(2\lambda) - 0.00028 z_{K_b})} \quad (3.7)$$

being K_b the number of measurement levels of the balloon and P_{K_b} and z_{K_b} the pressure and the height of the last level of balloon measurement respectively. The refraction index for each measurement level can be computed from the refractivity formula (Eq. 1.24), using the measurements of pressure, temperature and relative humidity (converted in water vapour partial pressure using Eqs. 1.13 and 1.14). The comparison with balloon measurements is depicted in Fig. 3.1. A correlation value of $\rho = 0.9102$ attests a good accuracy in the estimation of the zenith tropospheric delay from the GNSS station data.

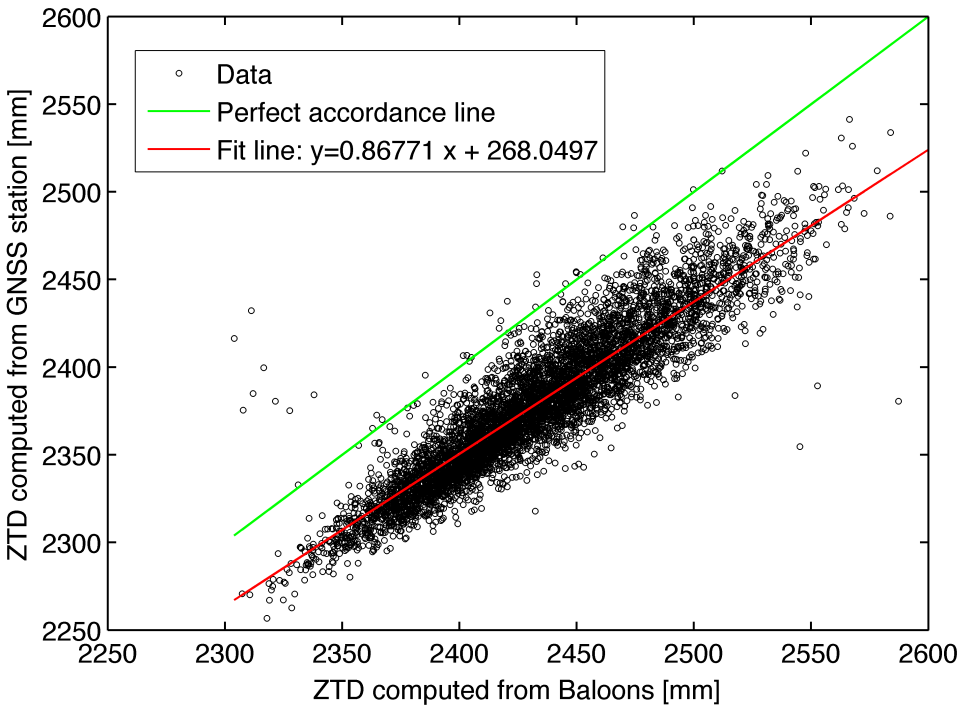


Figure 3.1: Comparison between GNSS-based and Balloon-based ZPD. Observation period is 2001-2010. Selected site is Cagliari. Both 00 and 12 Balloon data are included in the analysis.

The value of the RMSE is 56.955 mm. The deviation of the fit line with respect to the perfect matching line means that there is a relative bias within the domain of the variable values, superimposed on a visible dispersion greater for higher ZPD values. Knowing that the hydrostatic part of the delay is approximately 2300 mm and remains fairly constant, when atmospheric pressure is not excessively low, the variability of the ZPD depends mostly on the wet component. This dispersion may portend a worst

estimate of the content of water vapour, in presence of high atmospheric humidity. This actually occurs with the classic method of estimating the content of water vapour, since it is obtained directly from the ZPD, but not necessarily in the method proposed in this work, since it is based on a probabilistic approach, as we will better see later.

It should be noted that the approximation of the formula is valid for measures on the ground, while we have assumed the validity also in the highest level of balloon measures.

3.3 Computation of the precipitable water

The ZPD is available downstream the processing of GNSS receivers data. We thoroughly described that ZPD is given by an hydrostatic and a wet component (ZHD and ZWD respectively). If the GNSS receiver installation site is equipped with a station measuring pressure, the ZHD can be computed (§ 1.3.2) and the corresponding ZWD can be extracted (§ 1.3.3). In particular the flow of calculation first involves the application of the formula of Eq. 1.34, using the surface pressure measurement (P_0) latitude and height (lon, z_0) of the weather station, for the computation of ZHD. The second step is the conversion in IWV, by reversing the formula of Eq. 1.37:

$$IWV = \frac{ZWD \cdot M_w}{10^{-6} \cdot R} \left[K_2' + \frac{K_3}{T_m} \right]^{-1} \tag{3.8}$$

Beyond the constants (M_w, R) and coefficients (K_2', $K3$), which we thoroughly discussed § 1.3, the term T_m (as defined by Eq. 1.36) must be carefully chosen. An example to which we have referred in this work is provided by [37]. The weighted mean temperature (T_m) in Kelvin for the GNSS stations in the Mediterranean area can be approximated as a function of the surface temperature (T_0):

$$T_m = mT_0 + c \tag{3.9}$$

being $m = 0.7162$ and $c = 71.7054$ two empirical constants obtained by linear regression on a dataset of Italian balloon measurements.

Finally the precipitable water is obtained applying Eq. 1.38. The result is an estimation of the IWV or PW along the vertical column with good time resolution, theoretically up to a few seconds, but typical of 30 minutes, to have enough time for averaging. As reported in [38] several validation studies have compared the GNSS-IWV estimates with radiosoundings and microwave radiometer measurements. The errors associated with GNSS-IPW estimates as compared to radiosondes and Water Vapour Radiometers (WVR) are of the order of few millimeters.

We carried out a test on a ZPD one year dataset, we applied the PW estimation technique jointly using the surface measures of pressure (P_0) and temperature (T_0).

Fig. 3.2 shows a scatter diagram of the precipitable water estimated through the process just described, compared with that obtained by radiosondes measurements.

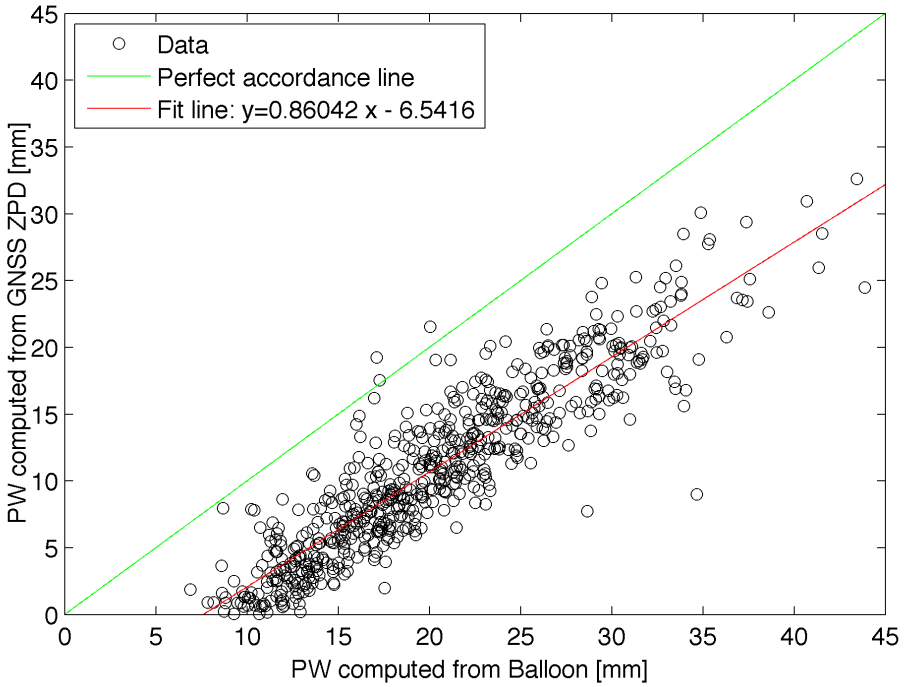


Figure 3.2: Comparison between GNSS-based and Balloon-based PW. Observation period is 2011 full year. The selected site is cagl. Both 00 and 12 Balloon data are included in the analysis.

The ZPD data are not the same of Fig. 3.1, but the results are consistent with what previously discussed, and therefore they give further confirmation. In fact the correlation is relatively high ($\rho = 0.9105$), and the value remains practically equal to that obtained for the ZPD. Even in this case there is a strong bias in the estimates of PW. The value of the angular coefficient of the straight line fit remains essentially unchanged, in this case too, the spread of values increases with the reference values, confirming what above described. In other words the relative error of the precipitable water content estimate is approximately constant varying precipitable water content. Finally the value of RMSE (5.0512 mm) is consistent with what is found in literature. In this example, only the results of a single test site are reported. Other considerations, and a more exhaustive framework will be made by analysing the results of all test sites chosen (see § 4.4 and § 5.2), comparing the standard technique described in the present section with the technique object of the present work.

3.4 Recent methods for 3D reconstruction of water vapour

The term Slant Water Vapour (SWV) refers to the full water vapour content along the satellite receiver line of sight. The values described above of PW are essentially

averages of all the available values of SWV over a time period ranging from 5 to 30 minutes, projected to the vertical column. The main advantage of averaging is the noise reduction, but it undoubtedly involves a partial loss of information. The availability of SWV estimates could help to reconstruct the tridimensional anisotropy of tropospheric water vapour content. For this reason in recent years a series of methods have been investigated and developed to extract the values of Slant Wet Delay (SWD) and consequently the SWV. This chapter does not want to do an exhaustive discussion of such practices but only to highlight their existence and describe some characteristics and prerogatives.

3.4.1 Slant wet delay retrieval

As the ZTD also SWD can be split in an hydrostatic (SHD) and a wet components (SWD). Most of techniques uses mapping functions to obtain the values of SWD and SHD (see Eq. 3.1):

$$STD = SHD + SWD = m_h(\theta^i_{ej}(t_j))ZHD(t_j) + m_w(\theta^i_{ej}(t_j))ZWD(t_j) \tag{3.10}$$

or alternatively introducing an asymmetry and the horizontal gradient as further unknown, as in Eq. 3.3. SWD or alternatively STD are the observables for the tomography and are directly connected to the wet refractivity (§ 1.3).

Most of software applications estimate the zenith hydrostatic component by using surface pressure measurements, and remove it from the observables, once they are projected to the zenith direction. The SWDs are given by mapping back to the line of tracking, the resulting zenith wet delays quantities and the horizontal gradients if computed. This can recreate the individual averaged slant delay measurements. The decomposition of the refractivity into a zenith component plus a horizontal gradient is certainly an approximation with consequent possible errors in the refractivity 3D structures reconstruction, and the loss of the ability to describe some features. However the transformation from the zenith occurs for two main reasons: to reduce the noise and to eliminate the hydrostatic component, otherwise difficult to eliminate.

The starting point of the tomography is a set of observables put in relation with the wet refractivity:

$$SWD = \mathbf{A}N_w \tag{3.11}$$

where $\mathbf{A}$ is the direct model matrix. This model must be inverted to reconstruct the 3D structure of the wet refractivity.

3.4.2 Tomographic approaches

The 4D-WV (time and space water vapour behavior) reconstruction has been recently addressed by the study and set up of tomography methods applied to GNSS observa-

tions. All methods must start from the estimate or the measurement of the STD or alternatively SWD.

In [39] a method is introduced for dividing the atmosphere in a number of volume cells (voxels), in each of which a value of refractivity N_k must be assigned. Each STD is given by the summation of each N_k intercepted by the signal ray, multiplied by the path length in that cell. Once the grid is defined, assuming the grid point refractivity values almost constant for an appropriate time window and given a sufficient number of observations from a dense network, the problem can be correctly addressed and solved using a least square estimation method. A method of this type provides an instantaneous estimate. To obtain the time evolution of the parameters some works have introduced the use of a Kalman filter, assuming a time variation model for refractivity (see for instance [40]).

Other methods are the Algebraic Reconstruction Techniques (ART), that process the $\mathbf{A}$ matrix by iterations without any inversion technique ([41]), processing row by row. An initial atmosphere status must be provided (e.g taken from a NWM run) in terms of atmosphere wet refractivity guess (N_w^0). For each iteration a new atmospheric state and the corresponding set of SWDs are computed (see Eq. 3.11). The values of the current estimate are compared with the measured ones. The result of this comparison is used to modify the current estimate, thereby creating a new estimate. The iteration stops when the desired accuracy is reached and the difference between the SWD^k at step k-th and the measured SWDs is below a certain threshold. The solution proposed by [41] uses the Multiplicative Algebraic Reconstruction Technique (MART) and can be summarised in the following formula:

$$\mathbf{x_j}^{k+1} = \mathbf{x_j}^{k} \cdot \left(\frac{SWD_i}{\mathbf{A_i} \cdot \mathbf{x^k}} \right)^{\frac{\lambda \mathbf{A_{ij}}}{|\mathbf{A_i}|^2}} \tag{3.12}$$

where $j = 1, ..., N$ denotes the grid cell, i the observation and k the iteration step. $\mathbf{x}$ is the value of wet refractivity, the $\mathbf{A^i}$ vector is the the row of the $\mathbf{A}$ matrix containing the elements of the ith observation in all the N grid cells, λ is the relaxation parameters.

Further improvements come from the use of phase measurements for the reconstruction of SWD. This kind of method ([42]) is completely different as introduces the calculation of the refractivity for each voxel directly in the receiver measurements data processing. Therefore the refractivity becomes a parameter to be estimated together with all others (clock errors, integer ambiguities etc ...). However, the availability of a sufficient number of satellites and receivers would lead a sufficient number of observations for a precise solution of the problem (although under-constrained), through the estimation of this high number of parameters.

Beyond the recent and ongoing developments almost all methods are characterised by some of the problems listed below:

- The number of observations is very important and must be sufficiently high to allow the solution of the problem. Due to the high temporal and spatial variability of water vapour, the tomography techniques require small and dense networks of GNSS stations, with consequent operating costs.

- The number of grid points must be chosen as a compromise between the desired resolution and the ability to reconstruct each voxel values by means of a sufficient number of observations. We must also take into account the computational costs of very high resolutions.

- The intersection of satellites rays is very irregular, both in time and space, as well as the accuracy of the estimate. This causes a non uniform distribution of observations, and a non uniform accuracy of the solutions.

- All conventional methods allow in some way the reconstruction of atmospheric wet refractivity. The transition to 3D water vapour content is not immediate since, according to Eq. 1.29, the temperature value must be known or estimated in every voxel. One possible solution is to use weather models either as ancillary data or as assimilation platform of the wet refractivity data.

3.5 Assimilation of ZPD in meteorological models

A Numerical Weather Predicion (NWP) model simulates the evolution of the atmospheric processes numerically solving a set of partial derivative equations of the fluid dynamics and thermodynamics that represent the physical processes of the atmosphere. Therefore they require initial and boundary conditions in order to operate and process data. The initial conditions of a model are a set of physical quantities defined typically through proper interpolations in all the grid points of the model and compatible with the model equations. They must be as accurate as possible, for an accurate data initialisation of the models.

Observations explore the reality through several instruments that provide informations about the real initial conditions to be ingested into the model. Unfortunately observations are few, sporadic and for localised points or areas and have their own uncertainty. For this reason they cannot be directly ingested by meteorological models and a preprocessing is needed to make them compatible with model input. The full process that elaborates the available observations and ancillary information to create the initial model conditions is called assimilation.

A state analysis is the initial state of the atmospheric variables to be assimilated into the model and is defined as a set of prognostic variables defined on all the grid points. This operatively derives from a combination of a background field distributed

on a regular grid and the observations which instead are spatially and temporally irregular.

Different types of initialisation techniques exist. In this chapter we will consider only a dynamic (nudging) and a variational (3D/4D-var) initialisation methods which are among the most used to assimilate ZPD/PW-based GNSS observations in numerical weather models.

The dynamic initialisation is conducted during a time period of the model equation integration on which the observations are used as constraints for the dynamics of the model variables. After this initialisation period the constraints are removed. The nudging technique falls into this category. During the initialisation time, the dynamic of a prognostic model variable (a) is driven by the model dynamical and physical equations ($F(a,t)$) adding an artificial forcing term, called nudging term, whose strength depends on the difference between the observed value and the coincident model field:

$$\frac{da}{dt} = F(a,t) + G(t)\sum_{i}^{N} w_i(a_i - a) \tag{3.13}$$

$G(t)$ is the nudging function generally increasing approaching the time the observations refer to; w_i is an analysis weight; a_i is one of the N observed values, and a is the model value. During the integration period the model output could be unrealistic, due to the nudging coefficient, that alters the physically-based structure of the model. After the initialisation time the nudging term is removed and the model dynamic is again governed by only physical equations. After an additional relaxation time (so called spin-up time) the outputs of the model will be considered as valid forecasts. As shown in [43] the nudging techniques has been applied for the assimilation of ground-based GNSS data, at Deutscher WetterDienst (DWD) and Meteo Swiss. An example of the nudging application is the use of precipitable water (PW^{obs}) estimated from ZPD as described in § 3.3 using, as ancillary data, pressure and temperature fields from the model. The specific humidity profile is computed for each model level (k) as:

$$q_v^{obs}(k) = \frac{PW^{obs}}{PW^{mod}} q_v^{mod}(k) \tag{3.14}$$

where the superscript ^{mod} denotes the model output of the variable. The so obtained specific humidity quantity constitutes the observation to be assimilated. A weighting function ($w_q(k)$) proportional to the thickness of the (k) level and to the saturated specific humidity is used in the nudging scheme (Eq 3.13).

The other method, that we want shortly describe, is the variational one, the most used in research and up to date also for the assimilation of GNSS-based observations into operational weather models. In the Three Dimension Variational (3D-var) assimilation process the aim is to identify the initial state as the most probable atmospheric state $\mathbf{x}$ at a given time, conditioned to the measurements $\tilde{\mathbf{y}}$ and to the forecast output

of the weather model $\tilde{\mathbf{x}}$ (at that time). This probability, applying the Bayes theorem, can be written as:

$$p(\mathbf{x}|\tilde{\mathbf{x}}\tilde{\mathbf{y}}) = \frac{p(\mathbf{x})p(\tilde{\mathbf{x}}\tilde{\mathbf{y}}|\mathbf{x})}{p(\tilde{\mathbf{x}}\tilde{\mathbf{y}})} \tag{3.15}$$

Assuming the worst case of the less informative a priori, i.e an uniform probability distribution for the prior atmospheric state variables ($p(\mathbf{x}) = const$) in the range of all possible values, considering $p(\tilde{\mathbf{x}}\tilde{\mathbf{y}})$ as a normalisation term constant with respect to $\mathbf{x}$ (when normalising $p(\mathbf{x}|\tilde{\mathbf{x}}\tilde{\mathbf{y}})$) and applying the product rule, the probability becomes:

$$p(\mathbf{x}|\tilde{\mathbf{x}}\tilde{\mathbf{y}}) = p(\tilde{\mathbf{x}}|\mathbf{x})p(\tilde{\mathbf{y}}|\mathbf{x}\tilde{\mathbf{x}}) \cdot const \tag{3.16}$$

In the second term only the conditioning on the atmospheric state $\mathbf{x}$ could be considered, since the knowing of the output model state ($\tilde{\mathbf{x}}$) is irrelevant for the measurement result ($\tilde{\mathbf{y}}$) when the actual atmospheric state ($\mathbf{x}$) is known. Moreover the dependence on the state variable $\mathbf{x}$ can be expressed also by introducing the observable operator $H(\mathbf{x})$, that maps from the model space to the observation space, such that $y = H(\mathbf{x})$:

$$p(\mathbf{x}|\tilde{\mathbf{x}}\tilde{\mathbf{y}}) = p(\tilde{\mathbf{x}}|\mathbf{x})p(\tilde{\mathbf{y}}|H(\mathbf{x})) \cdot const \tag{3.17}$$

The assimilation process must find the value of $\mathbf{x}$ that maximises the probability distribution on the left-hand term of Eq. 3.17. Assuming two Gaussians probability distribution function for the right-hand term of Eq. 3.17, i.e:

$$\begin{aligned} p(\tilde{\mathbf{x}}|\mathbf{x}) &= \frac{1}{\sqrt{2\pi B}} e^{\left[-\frac{1}{2}(\tilde{\mathbf{x}}-\mathbf{x})^T \mathbf{B^{-1}} (\tilde{\mathbf{x}}-\mathbf{x})\right]} \\ p(\tilde{\mathbf{y}}|H(\mathbf{x})) &= \frac{1}{\sqrt{2\pi R}} e^{\left[-\frac{1}{2}(\tilde{\mathbf{y}}-H(\mathbf{x}))^T \mathbf{R^{-1}} (\tilde{\mathbf{y}}-H(\mathbf{x}))\right]} \end{aligned} \tag{3.18}$$

the exponents of the two Gaussian functions must be minimised. $\mathbf{B}$ and $\mathbf{R}$ are the covariance matrices, of the NWP model status and the observations (including errors of representativeness) respectively:

$$\begin{aligned} \mathbf{B} &= \overline{(\tilde{\mathbf{x}} - \mathbf{x})(\tilde{\mathbf{x}} - \mathbf{x})^T} \\ \mathbf{R} &= \overline{(\tilde{\mathbf{y}} - H(\mathbf{x}))^T (\tilde{\mathbf{y}} - H(\mathbf{x}))} \end{aligned} \tag{3.19}$$

The upper bar indicates an average value over the samples number. In general the computation of $\mathbf{B}$ and $\mathbf{R}$ is difficult[2].

In the hypothesis of independence between uncertainties of the measurements and the model outputs, i.e. $(\mathbf{x} - \tilde{\mathbf{x}}) \cdot (\mathbf{y} - \tilde{\mathbf{y}})^T = 0$, the cost function ($J(\mathbf{x})$) to be minimised is:

$$J(\mathbf{x}) = \frac{1}{2}(\tilde{\mathbf{x}} - \mathbf{x})^T \mathbf{B^{-1}} (\tilde{\mathbf{x}} - \mathbf{x}) + \frac{1}{2}(\tilde{\mathbf{y}} - H(\mathbf{x}))^T \mathbf{R^{-1}} (\tilde{\mathbf{y}} - H(\mathbf{x})) \tag{3.20}$$

[2] A rigorous approach should require the knowledge of the true atmospheric state $\mathbf{x}$, that is obviously impossible. Some assumptions and proper techniques must be used to compute $\mathbf{B}$ and $\mathbf{R}$ (see for instance [44]).

The first term of the cost function is called background term because it is only model dependent. A further general difficulty is the inversion of the **B** and **R** matrices, due to their dimensions. The solution can vary case by case.

An example of 3D-var application for the initialisation of the fifth-generation of the NCAR Meososcal Model (MM5) is provided in [45]. A separate term in the cost function is introduced due to the GNSS observations (J_{GNSS}), defined as:

$$J_{GNSS}(\mathbf{x}') = \frac{1}{2}(\tilde{\mathbf{y}}'_{\mathbf{GNSS}} - H(\mathbf{x}'))^T \mathbf{R}^{-\mathbf{1}}_{\mathbf{GNSS}}(\tilde{\mathbf{y}}'_{\mathbf{GNSS}} - H(\mathbf{x}')) \tag{3.21}$$

where $\mathbf{x}'$ is the vector analysis increment such that $\mathbf{x_a} = \mathbf{x_b} + \mathbf{x}'$, $\tilde{\mathbf{y}}'_{\mathbf{GNSS}}$ is the ZTD observation, $\mathbf{x_b}$ is the background state vector, $\mathbf{x_a}$ the state analysis vector, $\mathbf{R_{GNSS}}$ the covariance matrix of the GNSS observation error. Assuming the observation errors uncorrelated, the **R** matrix becomes diagonal. The observation operator is the model simulation of the ZTD. It is directly given by applying Eqs. 1.34 and 1.37 on the model outputs. The final solution $\mathbf{x_a}$ is given by minimisating of the total cost function given by the background term and the conventional observation term of Eq. 3.20, adding the specific GNSS observation term (Eq. 3.21). Benefits of the ZTD assimilation have been found not only in the improvement of the forecast of the humidity content, but also in the decrease of errors in the wind components and temperature. The effects of assimilation of ZTD can be seen also in the precipitation forecast.

The assimilation of GPS ZTD data gives potential greater benefits using Four Dimension Variational (4D-Var) due to its ability to utilise the very high temporal resolution of the observations. The 4D-Var is an extension of 3D-Var where not only the instantaneous atmospheric state is under processing but also its step by step temporal evolution. The cost function in 4D-Var is a function of discrete increments of the atmospheric state. This is maximally exploited in the Rapid Update Cycle (RUC) operational weather prediction systems, running every hour out to at least 18h and with the consequent need of updated measurements data to perform the assimilation process. The same applies to the Rapid Refresh (RAP), the next-generation version of the 1-h cycle system that replaced the operational RUC at NOAA/NCEP.

Chapter 4

Probabilistic retrieval of atmospheric parameters

4.1 Theoretical aspects

In atmospheric science and physics the available measurements are in most of cases affected by relevant errors. These uncertainties should be considered through proper processes and models, to set up an inference process from incomplete and noisy information. In any experimental situations prior information can be available for any parameter involved in the experiment (this information is usually expressed in the form of a probability distribution). These may result from previous studies done on the same phenomenon or investigations or even inferred by the operator from some related knolegde. When this kind of additional information is wanted to be considered the Bayesian inference can be used to combine the prior information with the observations coming from experimental data, to obtain a posterior distribution. If the prior information is properly chosen and is representative of the involved experiment the posterior probability should represent clearly and precisely the phenomenon and the parameters.

In meteorology the prior information can be extracted from the analysis of time series of collected observations. Typically the choice of prior distribution is made between numerical weather model data or climatic weather observation data giving a well representativeness of the possible values occurrence. Particular attention should be made in the determination of the dataset extension, to find a compromise between the need to go back in time to have a large enough number of data, but not too much so as not to be affected by climate change.

4.1.1 The Bayes Theorem

According to probability theory, given an event B such that $Pr(B) > 0$[1] the conditional probability of the event A given the occurrence of the event B is defined as[2]:

$$Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)} = \frac{Pr(AB)}{Pr(B)} \tag{4.1}$$

In a dual manner the conditional probability of the event B given the occurrence of event A can be defined. By inversion of Eq. 4.1 it can be written:

$$Pr(AB) = Pr(A|B)Pr(B) = Pr(B|A)Pr(A) \tag{4.2}$$

Finally, by combination, the Bayes theorem can be obtained:

$$Pr(A|B) = \frac{Pr(B|A)Pr(A)}{Pr(B)} \tag{4.3}$$

Note that in Bayesian inference the most representative probability of an event A is not an intrinsic feature of the event, but it depends on the information available.

4.1.2 Application of the Bayes theorem to the measurement process

Let us define as x the true value of an observable and as $\tilde{x}$ the result of a measurement of x. In addition let us indicate with I is the background information containing all the additional knowledge (e.g. environmental conditions of the measurements).

The calibration of a measurement instrument is a process to obtain the so called likelihood probability distribution function $p(\tilde{x}|xI)$, which is the probability that given the true value x, the quantity $\tilde{x}$ is obtained as a result of the measure. Typically, the calibration is done in a controlled environment, as in a laboratory in which the value to be measured is known with very high accuracy.

On the other hand the goal is to find a way to lie the measurement result to the true value of the quantity. This is essentially equivalent to compute the probability distribution function: $p(x|\tilde{x}I)$. Through the Bayes theorem (Eq. 4.3) the calibration and the measurement probability distribution functions are related as:

$$p(x|\tilde{x}I) = \frac{p(x|I)p(\tilde{x}|xI)}{p(\tilde{x}|I)} \tag{4.4}$$

The value of $p(x|I)$ is the prior probability of x, i.e. all the information that can be drawn on the quantity to be measured, based on ancillary knowledge, previous measurements theoretical considerations, etc., for example the range of possible values for x, or better an estimation of the occurence of the x values in a certain period of the day or of the year, as obtainable from a proper dataset of x. The denominator term can be interpreted as a normalisation factor, which is normally calculated in such a way that $\int p(x|\tilde{x}I)dx = 1$ over all possible values.

[1] We indicate as Pr a probability, and as p a probability density.

[2] We indicate the expressions $Pr(AB) \equiv Pr(A \cap B)$ as the joint probability of events A and B.

4.1.3 Application of the Bayes theorem for a model inversion

4.1.3.1 Bayesian inference

A slightly different situation occurs when, for example using a model, the space of the observables[3] ($\mathbf{y}$) is not coincident with the space of the model states[4] ($\mathbf{x}$). In such a case an observable operator $f(\mathbf{x})$ is defined by the model as a function whose domain and codomain are the atmospheric states space and the observables space respectively.

In numerical models an important distinction is made between prognostic variables ($\mathbf{x}$) that appear in the equations of the model and the output variables ($\mathbf{y}$), defined diagnostic. The prognostic variables are used to reconstruct the link between the physical model variables at different time instants, by allowing the prediction to proceed forward in time. Therefore they constitute the state of the system. In numerical weather models diagnostic variables are calculated at every instant applying balance equations (mass and energy) and the equation of state of gases to prognostic variables. This set of equations is the operator of the model $f(\mathbf{x})$.

The model operator is an approximation of the reality that takes into account the knowledge of the physics of the involved variables. Consequently it is characterised by a specific accuracy $\boldsymbol{\epsilon}$:

$$\boldsymbol{\epsilon} = \mathbf{y} - f(\mathbf{x}) \tag{4.5}$$

If different and independent instruments are available to measure $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{y}}$ respectively, giving a sufficient number of observations, it is possible to compute the probability density functions: $p(\mathbf{x}|\tilde{\mathbf{x}}I)$ and $p(\mathbf{y}|\tilde{\mathbf{y}}I)$. Using the observable operator it is possible to compute $p(f(\mathbf{x})|f(\tilde{\mathbf{x}})I)$. Moreover for each pair of measurements $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$, relative to the same condition of the system under study, an indirect estimation of $\tilde{\boldsymbol{\epsilon}}$ can be extracted[5]. The probability distribution function $p(\boldsymbol{\epsilon}|\tilde{\mathbf{x}}\tilde{\mathbf{y}}I)$ is the conditional probability of each value of $\boldsymbol{\epsilon}$, the uncertainty of the model, given the real measures of the state and the observable, including both the instrumental uncertainty and the model uncertainty. It is generally difficult to have coincident measurements $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{y}}$; in addition the indirect measurement $\tilde{\boldsymbol{\epsilon}}$ is not strictly connected with $\boldsymbol{\epsilon}$, therefore in order to obtain this probability some additional considerations are required. For example as $\boldsymbol{\epsilon}$ is a not modeled quantity, it can be treated as a stochastic variable independent from the measured values and consequently $p(\boldsymbol{\epsilon}|I) = p(\boldsymbol{\epsilon}|\tilde{\mathbf{x}}\tilde{\mathbf{y}}I)$. However some assumptions and approximations must be introduced about the distribution form on the basis of observations and model parameters. Therefore it is difficult to obtain analytically the distribution $p(\boldsymbol{\epsilon}|I)$.

[3] An observable is a quantity measurable through proper measuring instruments. It can correspond to mathematical operators used in the calculation of measurable quantities.

[4] We introduce the vector notation as in general we will consider observable and model output vectors as composed by different parameters.

[5] Let us indicate with $\tilde{\epsilon}$ a measurement of the model operator error (see Eq. 4.5).

4.1.3.2 Model Inversion

The aim of the inversion of a model is to find the probability distribution $p(\mathbf{x}|\tilde{\mathbf{y}}I)$. Using the Bayes theorem it can be written:

$$p(\mathbf{x}|\tilde{\mathbf{y}}I) = \frac{p(\mathbf{x}|I)p(\tilde{\mathbf{y}}|\mathbf{x}I)}{p(\tilde{\mathbf{y}}|I)} \tag{4.6}$$

where the $p(\mathbf{x}|I)$ is the prior distribution of the available dataset of vector states and the denominator term $p(\tilde{\mathbf{y}}|I)$ can be treated as a normalisation factor.

The remaining term, using the marginalisation rule can be written as:

$$p(\tilde{\mathbf{y}}|\mathbf{x}I) = \int_{-\infty}^{+\infty} p(\tilde{\mathbf{y}}\boldsymbol{\epsilon}|\mathbf{x}I)d\boldsymbol{\epsilon} \tag{4.7}$$

By applying the product rule:

$$p(\tilde{\mathbf{y}}\boldsymbol{\epsilon}|\mathbf{x}I) = p(\boldsymbol{\epsilon}|\mathbf{x}I)p(\tilde{\mathbf{y}}|\boldsymbol{\epsilon}\mathbf{x}I) \tag{4.8}$$

The probability distribution $p(\boldsymbol{\epsilon}|\mathbf{x}I)$ is result of the verification of the model and can be written as:

$$p(\boldsymbol{\epsilon}|\mathbf{x}I) = p(\mathbf{y} - f(\mathbf{x})|\mathbf{x}I) \tag{4.9}$$

If $\boldsymbol{\epsilon}$ and $\mathbf{x}$ are given is also given $\mathbf{y}$ using Eq. 4.5, so it can be written:

$$p(\tilde{\mathbf{y}}|\boldsymbol{\epsilon}\mathbf{x}I) = p(\tilde{\mathbf{y}}|\mathbf{y}I) \tag{4.10}$$

where the last distribution is given by the calibration of the instrument. Moreover if $\mathbf{x}$ is given, it is a constant quantity, as well as $f(\mathbf{x})$, and thus from Eq. 4.5: $d\boldsymbol{\epsilon} = d\mathbf{y}$. The integral of Eq. 4.7 can be rewritten as:

$$p(\tilde{\mathbf{y}}|\mathbf{x}I) = \int_{-\infty}^{+\infty} p(\tilde{\mathbf{y}}\boldsymbol{\epsilon}|\mathbf{x}I)d\boldsymbol{\epsilon} = \int_{-\infty}^{+\infty} p(\mathbf{y} - f(\mathbf{x})|\mathbf{x}I)p(\tilde{\mathbf{y}}|\mathbf{y}I)d\mathbf{y} \tag{4.11}$$

Finally the posterior probability of Eq. 4.6, considering the quantity $p(\tilde{\mathbf{y}}|I)$ a normalisation factor, can be rewritten as:

$$p(\mathbf{x}|\tilde{\mathbf{y}}I) \propto p(\mathbf{x}|I) \int_{-\infty}^{+\infty} p(\mathbf{y} - f(\mathbf{x})|\mathbf{x}I)p(\tilde{\mathbf{y}}|\mathbf{y}I)d\mathbf{y} \tag{4.12}$$

Therefore it is necessary to know the conditional probability of each possible real observable value $\mathbf{y}$ given model state $\mathbf{x}$ to compute the integral of Eq. 4.12 and to obtain the posterior probability of the model state. Generally this problem can not be solved analytically, unless some assumptions are made. The definition of the likelihood of Eq. 4.11 requires specifying the model error distribution. Further development of Eq. 4.12 can be made if several measurement $\tilde{\mathbf{y}}$ are available and the measurement uncertainties are known. Variance and covariance parameters can be included in the inversion scheme for the assessment of the probability distribution of the model uncertainty given the model state.

4.1.3.3 The adopted scheme

A completely different situation occurs when there is no way to define or to estimate the conditional probability $p(\boldsymbol{\epsilon}|\mathbf{x}I)$. In fact in many cases it is not possible to estimate $\boldsymbol{\epsilon}$ given the state of the model, because normally it is not possible to separate the observable measurement error from the model error. In such cases Eq. 4.12 can not be applied and it is necessary to find another method to compute the likelihood distribution $p(\tilde{\mathbf{y}}|\mathbf{x}I)$.

If, as in the case of the present work, there is the availability of a large dataset of independent measurements and state vectors (for example the output of a model), it can be exploited to find out any eventual correlation between observables measurements and status variables. An attempt to find out an empirical or semi-empirical relationship able to provide the likelihood probability distribution can be consequently made.

The aim of this work is to derive an estimate of the atmospheric state starting from a series of ground measurements. The vertical atmospheric state is determined if the values of the L atmospheric variables are known in the vertical column (namely the vertical profiles of $L = 3$ variables: pressure, temperature and humidity, as seen in § 1). A vector state $\mathbf{x}$ can be intended as a discretisation of the atmospheric state by sampling it at N different levels above the Earth surface.

A representative set of some physical observables $\mathbf{y}$ must be chosen, whose measures may allow to express the likelihood probability distribution $p(\tilde{\mathbf{y}}|\mathbf{x}I)$. Once those physical observables are measured, the probability distribution for the possible $N \times L$ components of the atmospheric state vectors $\mathbf{x}$ can be obtained using Eq. 4.6.

Plausible expressions for the $p(\mathbf{x}|I)$ prior may be obtained using a large set $\mathbf{x_i},\ \mathbf{i} = 1, ...M$ of possible atmospheric states, thought representative of the meteorological situation under observation. In practice a suitable set can be generated by a physical atmospheric model, as done for the ensemble forecasts, or extracted from a database of real measurements, or as a combination of them. The model inversion equation (Eq. 4.6) leads to the discrete posterior probability of each atmospheric state:

$$Pr(\mathbf{x_i}|\tilde{\mathbf{y}}I) = Pr(\mathbf{x_i}|I)p(\tilde{\mathbf{y}}|\mathbf{x_i}I)B \tag{4.13}$$

where B is a normalisation constant including the prior $p(\tilde{\mathbf{y}}|I)$. It can be calculated simply imposing the normalisation of the probability $Pr(\mathbf{x_i}|\tilde{\mathbf{y}}I)$ all over the M possible states of the set $\mathbf{x_i}$. For instance, if the same prior probability $1/M$ is assigned to each state (as in our case), we have explicitly:

$$B = \frac{M}{\sum_{i=1}^{M} p(\tilde{\mathbf{y}}|\mathbf{x_i}I)} \tag{4.14}$$

This approach is oriented to data analysis and, therefore, must be customised to the application. In this work this scheme is adopted and some application results are presented. The following is a description of the data, their relationships, the step made for the approximation of the likelihood probability distribution and the results obtained by adopting this scheme.

4.2 Description of datasets

For an application of the method and to gain some insight on its effectiveness, as set $\mathbf{x_i}$ of possible atmospheric states we have taken the analysed fields of Modern Era Retrospective-analysis for Research and Applications (MERRA), an atmospheric reanalysis model designed and developed by NASA. It is used for research and applications and is equipped with the Goddard Earth Observing System Model, Version 5 Data Assimilation Scheme (GEOS-5 DAS), a state-of-the-art GEOS-5 data assimilation scheme, developed by the Global Modeling and Assimilation Office (GMAO) allowing the ingestion of many data from modern observing systems. The GEOS-5 DAS is designed to integrate available observing data, including both satellite and ground-based measurements.

As reanalyses are ideal for investigating the climatological conditions, they are reliable for the characterisation of the prior probability distribution $p(\mathbf{x_i}|I)$. The available MERRA time series spans over a long term period, from 1979 to present. This dataset consists of four simulations per day at synoptic times (00, 06, 12 and 18 UTC) of three dimensional atmospheric analyses over the model spatial grid resolution of 1/2 degrees latitude and 2/3 degrees longitude. Several products in several configuration can be chosen and selected.

The product used in the present analysis is named: $inst6\_3d\_ana\_Np$; it stands for MERRA instantaneous 6 hourly 3D analysis at N pressure model levels. Data are distributed using NetCDF file format. Available variables are listed in Table 4.1.

Variable	**Dimensions**	**Description**	**Units**
SLP	2D	Sea Level Pressure	Pa
PS	2D	Surface Pressure	Pa
h	3D	Geopotential Height	m
T	3D	Air Temperature	K
u	3D	Eastward wind component	m s^{-1}
v	3D	Northward wind component	m s^{-1}
q	3D	Specific humidity	Kg Kg^{-1}
$O3$	3D	Ozone mixing ratio	kg kg^{-1}

Table 4.1: Volume concentration of the major gaseous constituents of the atmosphere for dry air ([46]).

They are instantaneous model analysis fields resulting from the Grid-point Statis-

tical Interpolation (GSI) scheme, produced on the full resolution native horizontal grid and on 42 vertical pressure levels listed in Table 4.2.

1000 hPa	975 hPa	950 hPa	925 hPa	900 hPa	875 hPa	850 hPa
825 hPa	800 hPa	775 hPa	750 hPa	725 hPa	700 hPa	650 hPa
600 hPa	550 hPa	500 hPa	450 hPa	400 hPa	350 hPa	300 hPa
250 hPa	200 hPa	150 hPa	100 hPa	70 hPa	50 hPa	40 hPa
30 hPa	20 hPa	10 hPa	7 hPa	5 hPa	4 hPa	3 hPa
2 hPa	1 hPa	0.7 hPa	0.5 hPa	0.4 hPa	0.3 hPa	0.1 hPa

Table 4.2: 42 Pressure levels of MERRA output data ([46]).

The dataset has been divided by season and time of the day (sixteen subsets on the whole). This allows a fuller exploitation of the prior information on both the season and the time of the observations, which clearly affect the a priori expected values, in particular the amount of humidity and temperature. The collected dataset cover the grid area shown in Fig. 4.1 and the time span is a ten years period (2001 - 2010).

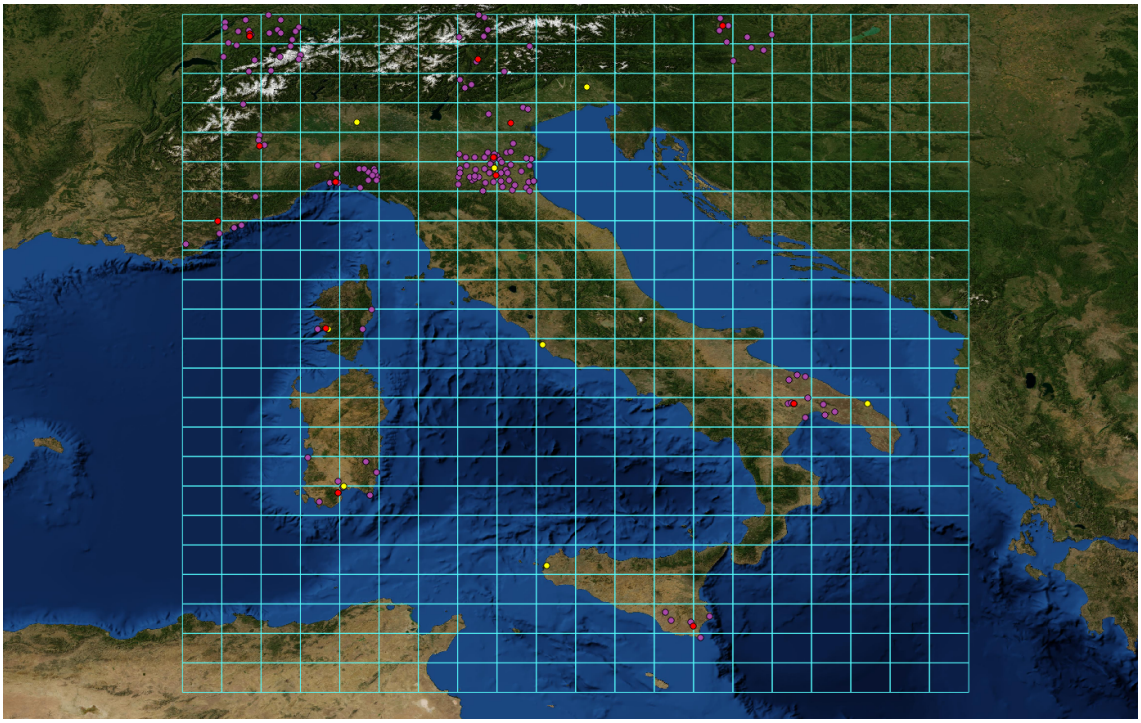


Figure 4.1: View of the distribution of IGS GNSS stations (red points), baloon launch sites (yellow points) and MERRA model pixels (cyan rectangles) over the mediterranean area. The weather stations in the four MERRA pixel nearest the IGS stations are also shown (pink points). The background Map of the mediterranean area has been obtained from the Blue Marble Next Generation image series ([47]).

As concerns the physical quantities, the MERRA values of ground pressure (P_0) and geopotential height (h), temperature (T) and specific humidity (q) at 42 isobaric levels provide a reasonable description of the atmosphere state. Since the specific humidity depends directly on pressure (see Eq. 1.11), it has been converted into WV partial pressure e_w, an intensive independent quantity like T and h. Ultimately any

atmospheric state is assumed represented by the 127 components vector:

$$\mathbf{x} = [PS,\ h_1,\ T_1,\ e_{w1},\ h_2,\ T_2,\ e_{w2},\ ...,\ h_{42},\ T_{42},\ e_{w42}] \tag{4.15}$$

The whole dataset consists of a maximum of more than 900 different atmospheric states for each grid pixel, for each season and for each time slot.

The observable vector has been chosen selecting an opportune set of ground-based measurable quantities, providing information about the ground layer status and able to construct a robust and reliable likelihood distribution. The observable vector is:

$$\mathbf{y} = [P_0,\ T_0,\ e_{w0},\ ZPD_{GNSS}] \tag{4.16}$$

It is measured by suitable weather station instruments located at the ground level, where a GNSS receiver for the tropospheric delay detection (ZPD_{GNSS}) can be also placed. As further information the heights of the weather station and the GNSS receiver locations must be considered.

Although the same notation is adopted, rigorously the ground level observable P_0 is not the same physical quantity provided by MERRA data (PS), since the latter represents a mean value over a wide area, and is relative to a mean ground level height generally different from that of the ground instruments. Anyway the pressure field exhibits a weak spatial variation and the locally measured values can be considered fully representative of the mean value over horizontal distances of a few tenth of kilometers (at the constant reference height).

Jointly to the MERRA reanalysis data, real measurements carried out by weather balloons on a 11 years period (2001-2010) have been used. Baloon dataset have been obtained from the Department of Atmospheric Science of the University of Wyoming ([48]), that archives all the balloon soundings launched in the planet. The balloon data are indicated with yellow dots in Fig. 4.1.

Moreover the red dots of the map are the installation sites of the IGS reference station available in the area of interest. In particular the five minutes sampled ZTD data processed by JPL with the Gipsy-Oasis software were used for our application (see § 3.2 for more details). The IGS stations dataset has been constructed over the same 11 years period.

The measurement data from the weather stations located in the 4 neighbor MERRA pixels around the GNSS station site have been also collected for the full time period (2001-2011). Data have been made available from the database of LAMMA Consortium ([49]), a database dedicated to the collection of measured data for the observation of weather phenomena taking place in the Mediterranean, functional to the operational activities for civil protection purposes in cooperation with the Centro Funzionale Regionale of Tuscany Region ([50])[6].

[6]The source of the database used in this work have been: Agenzia Regionale per la Preven-

In this work the retrieval developed technique results are shown exclusively for the sites in which they are available balloon stations launch for their verification. Therefore, the GNSS stations (with the corresponding meteorological stations and pixels of the MERRA model) to which we refer are Ajaccio (ajac), located in the central part of the Mediterranean Sea in the west coast of Corsica, Cagliari (cagl) located in the southern part of the Sardinia and finally Medicina (medi) located in the Po valley. This set of stations have been used to verify the algorithm in areas with different weather and climate: near the sea, in the proximity of mountains, and in an area often subject to temperature inversions and fog formation. It should be noted that the technique is applicable in all the MERRA grid points in which there are simultaneously available a measure of ZPD and the measurements of pressure temperature and humidity at the surface level.

4.3 Algorithm implementation

As mentioned in § 4.1.3.3, once the components of the status vectors $\mathbf{x}$ and the observable vector $\mathbf{y}$ have been defined, it is necessary to find out a relationship for the representation of the conditional probability distribution function $p(\tilde{\mathbf{y}}|\mathbf{x}I)$, to apply the Bayes theorem in the form of Eq. 4.6 discretised in Eq. 4.13. Reasonably, as explained below, some relationship between the atmospheric state $\mathbf{x}$ and the measured observables can be assumed. Doing so an algorithm for assigning a posterior probability to any state $\mathbf{x_i}$ can be obtained, once the measurement results ($\tilde{\mathbf{y}}$) are given. For the explicit calculation of the distribution $p(\tilde{\mathbf{y}}|\mathbf{x}I)$ we can apply the product rule to the expression for $\tilde{\mathbf{y}}$ given by the observable vector (Eq. 4.16), obtaining:

$$p(\tilde{\mathbf{y}}|\mathbf{x}) = p(\tilde{P}_0|\mathbf{x}) \cdot p(\tilde{T}_0|\tilde{P}_0\mathbf{x}) \cdot p(\tilde{e}_{w0}|\tilde{P}_0\tilde{T}_0\mathbf{x}) \cdot p(Z\tilde{P}D_{GNSS}|\tilde{P}_0\tilde{T}_0\tilde{e}_{w0}\mathbf{x}) \tag{4.17}$$

where the generic ancillary information I has been omitted for brevity. The development of Eq. 4.17 presents a hierarchy chosen on the basis of the correlations between the variables that have been analysed in a preliminary step and successively also used to extrapolate some relationship between the variables.

The measured ground pressure $\tilde{P}_0$ depends mainly on the state variable P_0 included in $\mathbf{x}$ and a linear relation between them can be reasonably imposed as:

$$\tilde{P}_0 = \alpha_P + \beta_P P_0 \tag{4.18}$$

zione e Protezione Ambientale dell'Emilia Romagna - Servizio Meteorologico Regionale (ARPA-SMR)([51],[52]) , Ufficio Idrografico e Mareografico di Bologna, Ufficio Idrografico e Mareografico di Parma, Dipartimento di Ingegneria delle Costruzioni dell'Ambiente e del Territorio - Universita' di Genova ([53]), Global Telecommunication System, ITAV - Aeronautica militare, Sistema Informativo Agricolo Nazionale ([54]).

The measured ground temperature $\tilde{T}_0$ depends mainly on the temperature T1 of the nearest (to ground) level and the measured ground pressure, again in a linear way:

$$\tilde{T}_0 = \alpha_T + \beta_T T_1 + \gamma_T \tilde{P}_0 \tag{4.19}$$

The measured ground WV partial pressure, $\tilde{e}_{w0}$, depends linearly on the WV partial pressure, e_{w1}, of the nearest (to ground) level as well as on the measured ground temperature and pressure, i.e.:

$$\tilde{e}_{w0} = \alpha_e + \beta_e e_{w1} + \gamma_e \tilde{T}_0 + \eta_e \tilde{P}_0 \tag{4.20}$$

The measured tropospheric zenith delay $\tilde{ZPD}_{GNSS}$ is modeled by the delay given by the state $\mathbf{x}$ starting from the first level up o the satellite height in the vertical direction, plus the delay between ground and the first level itself. The former is calculated as the integral of the atmospheric refractivity $n-1$ along the vertical coordinate z, taking advantage of the Thayer's semi-empirical formula (see for instance [55]) reported in this work on Eq. 1.27, and opportunely converted in:

$$n - 1 = \left(\frac{K_1 R}{M_d}\right) \times 10^{-6} \cdot \rho_T + \left(K_2' Z_w^{-1}\right) \times 10^{-6} \cdot \frac{e_w}{T} + \left(K_3 Z_w^{-1}\right) \times 10^{-6} \cdot \frac{e_w}{T^2} \tag{4.21}$$

The measured wet delay, i.e. the total GNSS delay minus the hydrostatic contribution, is so assumed linearly dependent on the MERRA wet delay (computed from level 1 up to 42), with the remaining wet contribution of the ground layer proportional to the first level distance from ground. Explicitly:

$$\begin{aligned} \tilde{ZPD}_{GNSS} - \frac{c_1 \tilde{P}_0}{g_{eff}} &= \alpha_{ZPD} + \beta_{ZPD} \left(ZPD_{MERRA}(\mathbf{x}) - c_1 \frac{P_1 - P_{42}}{g_{eff}} \right) + \\ &+ \gamma_{ZPD} \left(\left\langle c_2 \frac{e_w}{T} \right\rangle + c_3 \left\langle \frac{e_w}{T^2} \right\rangle \right) (z_1 - z_0) \end{aligned} \tag{4.22}$$

where g_{eff} is the effective gravity (see § 1.3.2) computed using Eq. 1.32, z_0 is the height of the weather station position, z_1 is the height of the first isobaric level computed from the geopotential height h_1 using Eq. 1.5 and the constant c_1,c_2,c_3, are respectively (see § 1.3.1):
$c_1 = \frac{K_1 R}{M_d} \times 10^{-6} = 222.77 \times 10^{-6}$ m^3 kg^{-1} $\equiv 222.77 \times 10^{-4}$ m^2 s^{-2} hPa^{-1};
$c_2 = K_2' \times 10^{-6} = 16.5 \times 10^{-6}$ K hPa^{-1};
$c_3 = K_3 \times 10^{-6} = 3.776 \times 10^{-1}$ K^2 hPa^{-1};
the $\langle \dots \rangle$ denotes the mean between the ground and the first level values of the quantity enclosed in the angle brackets; explicitly:

$$\begin{aligned} \left\langle \frac{e_w}{T} \right\rangle &= \frac{\left(\frac{\tilde{e_{w0}}}{\tilde{T}_0} + \frac{e_{w1}}{T_1} \right)}{2} \\ \left\langle \frac{e_w}{T^2} \right\rangle &= \frac{\left(\frac{\tilde{e_{w0}}}{\tilde{T}_0^2} + \frac{e_{w1}}{T_1^2} \right)}{2} \end{aligned} \tag{4.23}$$

Moreover the value of $ZPD_{MERRA}(\mathbf{x})$ is computed as a discrete piecewise linear integration of the refraction index:

$$ZPD_{MERRA}(\mathbf{x}) = \sum_{l=1}^{41} \left(\frac{\mathbf{n_l} + \mathbf{n_{l+1}}}{2} - 1 \right) (\mathbf{z_{l+1}} - \mathbf{z_l}) \tag{4.24}$$

where $\mathbf{n_i}$ is the refraction index at the **i**-th MERRA isobaric level computed applying Eqs. 1.24 and 1.23 to values of P_i,T_i, and e_{wi} ($P_{di} = P_i - e_{wi}$). In c_2 and c_3 coefficients the inverse compressibility factor of water vapour (Z_w^{-1}) has been considered equal to one, and consequently omitted. In effect this assumption does not introduce a significant error. As can be seen on Fig. 4.2 the compressibility factor values computed on MERRA model data using Eq. 1.25 are not very different from one and the error associated to this approximation produces effects to the maximum of a millimeter in the total delay computation, which falls within the range of fluctuations for measurements and calculated parameters.

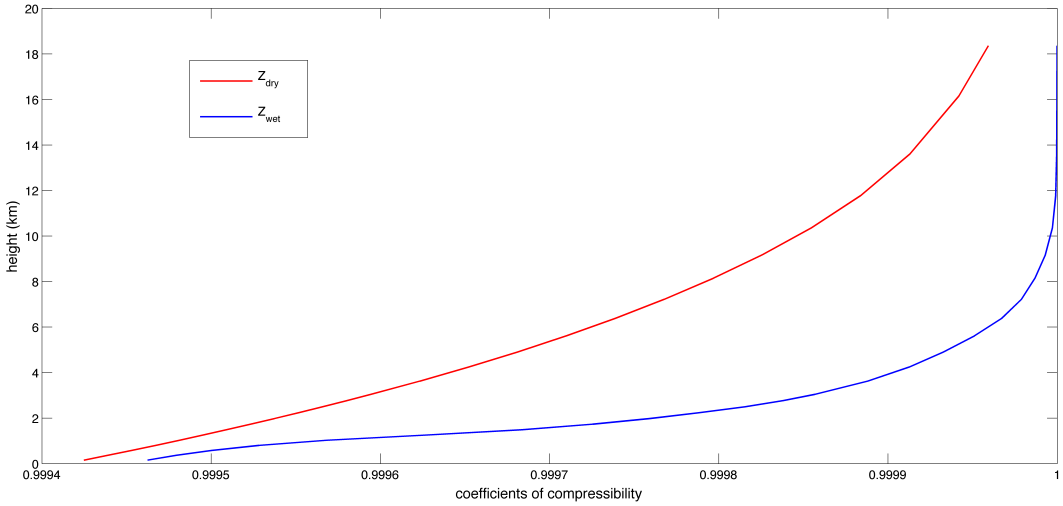


Figure 4.2: Vales of compressibility factors computed applying Eq. 1.25 on MERRA model data. The plotted values are obtained by averaging the MERRA data profiles of T, P and e relative to the Cagliari pixel on winter values at 12:00 UTC over the ten years period (2001-2010).

More in detail the left hand side of the equation Eq. 4.22 contains the ZWD, which is given by the difference between the measured total tropospheric delay and the hydrostatic component computed as a function of the surface pressure (see § 1.3.1). The right hand side is consequently modeled: as the MERRA model gives only the vertical values of pressure, temperature and specific humidity starting from the first isobaric level (P_1 = 1000 hPa) which generally is not coincident with the surface pressure the, difference between the total delay and the hydrostatic component computed on MERRA profile (second adding term) must be corrected for the gap of the zenith wet quantity between the surface level and P_1 (third adding term).

Unlike the three previous relationships, Eq. 4.22 is not simply an empirical formula to realign the model values with those locally measured, but also has a physical content.

If the MERRA atmospheric states are a good approximation of the actual situation determining the measurements results, then the physical meaning of each single term of Eq. 4.22 is expected to be retained even after the adjustment of the parameters. Moreover the fitting procedure can correct eventual biases in the MERRA data. In particular the values for β_{ZPD}, which ensure the best realignment, should be positive and possibly close to unity (being the factor of the largest term containing only reanalysis quantities), with an absolute value for the bias α_{ZPD} at most of the same order of the measured wet delay ($\simeq 10$ cm).

All the realignment parameters entering these models are determined by best fit procedures on the set of available MERRA data and the corresponding ground observations. As can be noticed linear dependencies have been assumed, in decreasing order of the correlation strength. Thus the maximum correlation value has been exploited for the linear model involving only one variable, then decreasing the correlation values, the number of involved variables of the linear model has been increased, and so on. These linear dependencies can be easily identified from the analysis of scatter plots of the chosen test sites showing the correlation between the measured quantities and the status variables of Eqs. 4.18 – 4.21 (see Figs. 4.3 – 4.5).

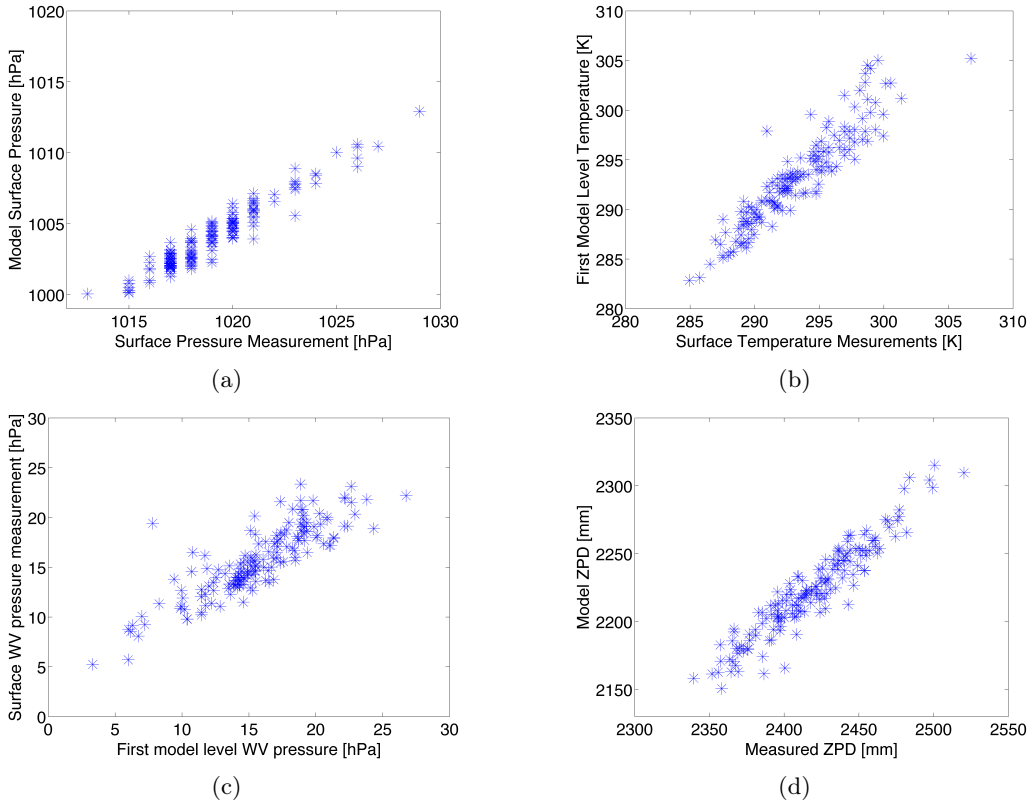


Figure 4.3: Scatter plots of model variables vs. measured parameters. Period is spring at 12:00 UTC. Station is ajac (Ajaccio).

Such preliminary data analysis has been made on three chosen test sites, which the results of the present work refer to.

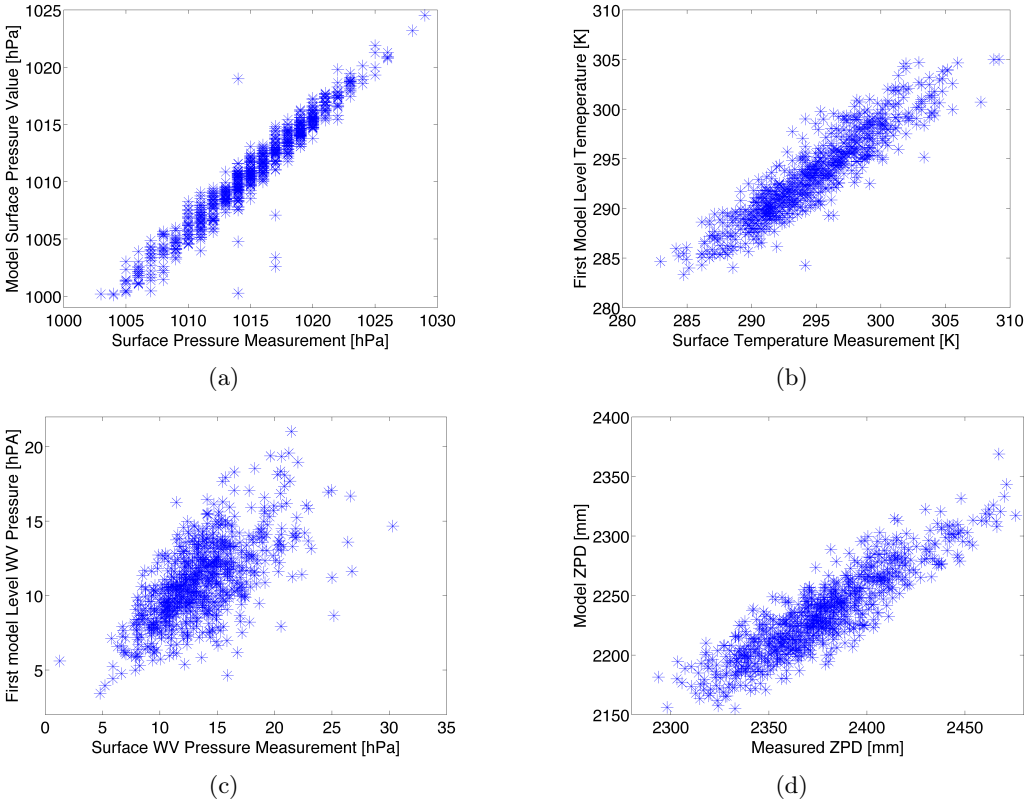


Figure 4.4: Scatter plots of model variables vs. measured parameters. Period is spring at 12:00 UTC. Station is cagl (Cagliari).

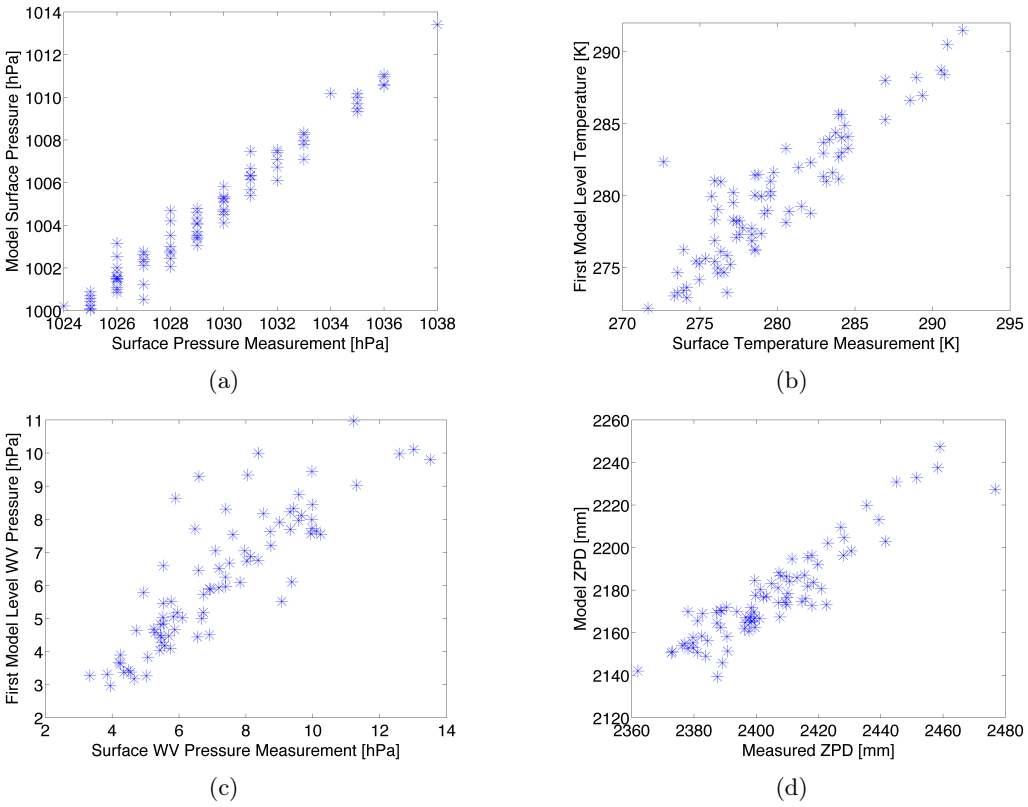


Figure 4.5: Scatter plots of model variables vs. measured parameters. Period is winter at 12:00 UTC. Station is medi (Medicina).

It should be noted that the scatter plot of the pressure values presents a series of horizontal empty spaces due to the measurement accuracy of the used barometric station limited to one hPa. Strong corelations are shown by plots, particularly for the pressure, the temperature and the tropospheric delay. A good correlation is also for the water vapour content, although with a very high spread of values. This is not surprising if we consider that the used model relates the water vapour partial pressure measured at ground level by the weather station, with the water vapour pressure estimated by the model at the first level (1000 hPa). Indeed the height of the first level of MERRA data varies depending on the ground pressure value and the of water vapour content on the vertical can significantly vary even in few hundred meters.

The leading correlations of the variables in Eqs. 4.18 – 4.22 allow us to make with good approximation the assumptions:

$$\begin{aligned} p(\tilde{P}_0|\mathbf{x}) &\simeq p(\tilde{P}_0|P_0) \\ p(\tilde{T}_0|\tilde{P}_0\mathbf{x}) &\simeq p(\tilde{T}_0|P_0T_1) \\ p(\tilde{e}_{w0}|\tilde{P}_0\tilde{T}_0\mathbf{x}) &\simeq p(\tilde{e_{w0}}|\tilde{P}_0\tilde{T}_0e_{w1}) \\ p(Z\tilde{P}D_{GNSS}|\tilde{P}_0\tilde{T}_0\tilde{e}_{w0}\mathbf{x}) &\simeq p(Z\tilde{P}D_{GNSS}|\tilde{P}_0\tilde{T}_0\tilde{e}_{w0}ZPD_{MERRA}(\mathbf{x})h_1T_1e_{w1}g_{eff}) \end{aligned} \tag{4.25}$$

In other words the observable vector measurements are assumed as dependent only on the state variables shown in the model. All the other dependencies of the measurements are assumed as stochastic, and consequently processed as stochastic variables (independent from the components of the state vector $\mathbf{x}$). Therefore once the best estimates for all the model parameters entering Eqs. 4.18, – 4.22 are found, the probability distributions in the factorisation coincide with those of the residuals. As can be assessed from the residual histograms, these probability distributions are well represented by normal distributions centreed on the model values and variances given by the mean square residuals. We can examine it in the following figures (Figs. 4.6 – 4.8), which show the normalised histograms of the residual values given from the differences between the measured values and those obtained from the above mentioned equations (Eqs. 4.18 – 4.21), with the parameters computed by the fit on data represented in Figs. 4.3 – 4.5.

It is evident that the normal distribution well approximates the residual distributions when the number of observations is sufficient to give an enough statistics, as in the case of the Cagliari case (Fig. 4.7). When the number of observations is very low, as seen in the case of Medicina (Fig. 4.8), the histogram start deviating from the distribution, which, however, seems to be still well representative of the actual distribution.

As a consequence given a set of measurements $\tilde{\mathbf{y}}$ the probability distribution $p(\tilde{\mathbf{y}}|\mathbf{x})$

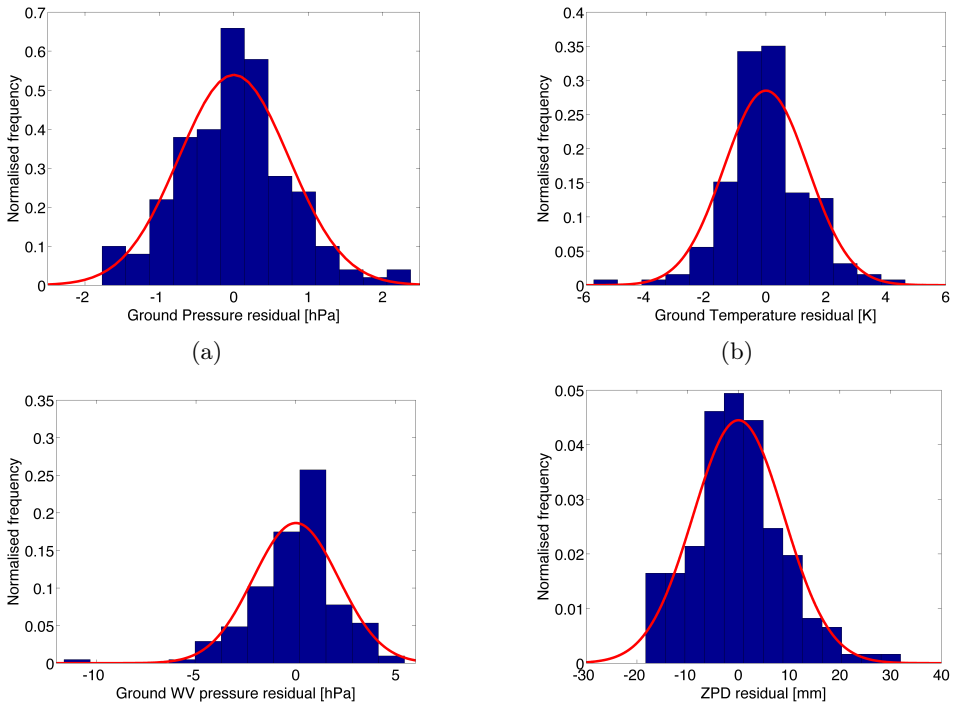


Figure 4.6: Distributions of residual of model quantities minus measured parameters (blue normalised histograms). In red are reported the resulting normal distributions using the variances values. Period is spring at 12:00 UTC. Station is ajac (Ajaccio).

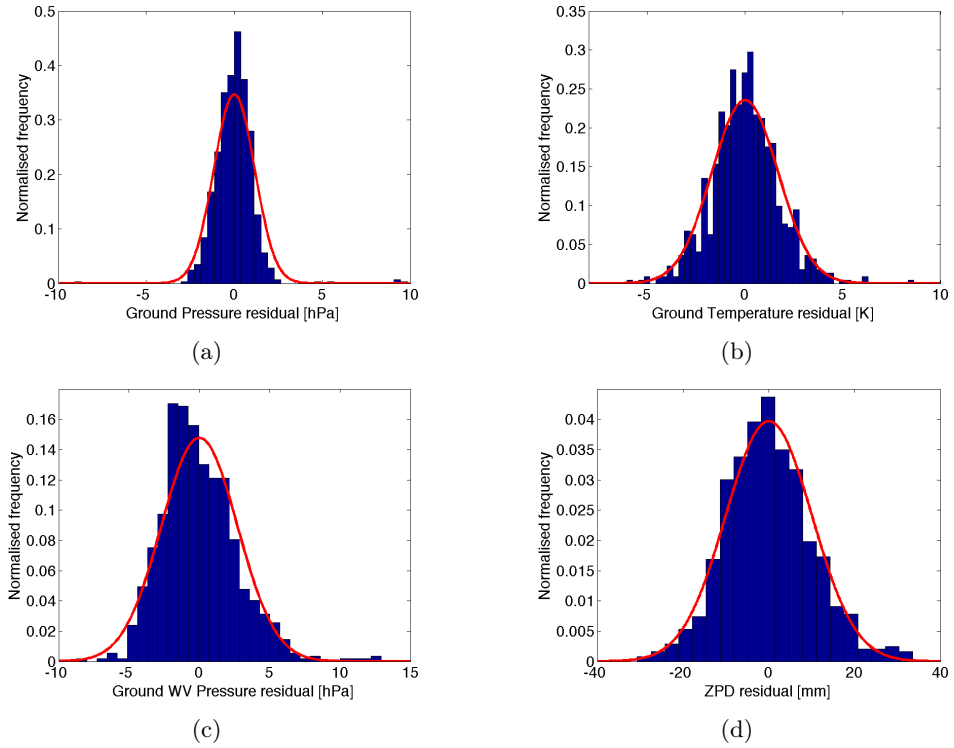


Figure 4.7: Distributions of residual of model quantities minus measured parameters (blue normalised histograms). In red are reported the resulting normal distributions using the variances values. Period is spring at 12:00 UTC. Station is cagl (Cagliari).

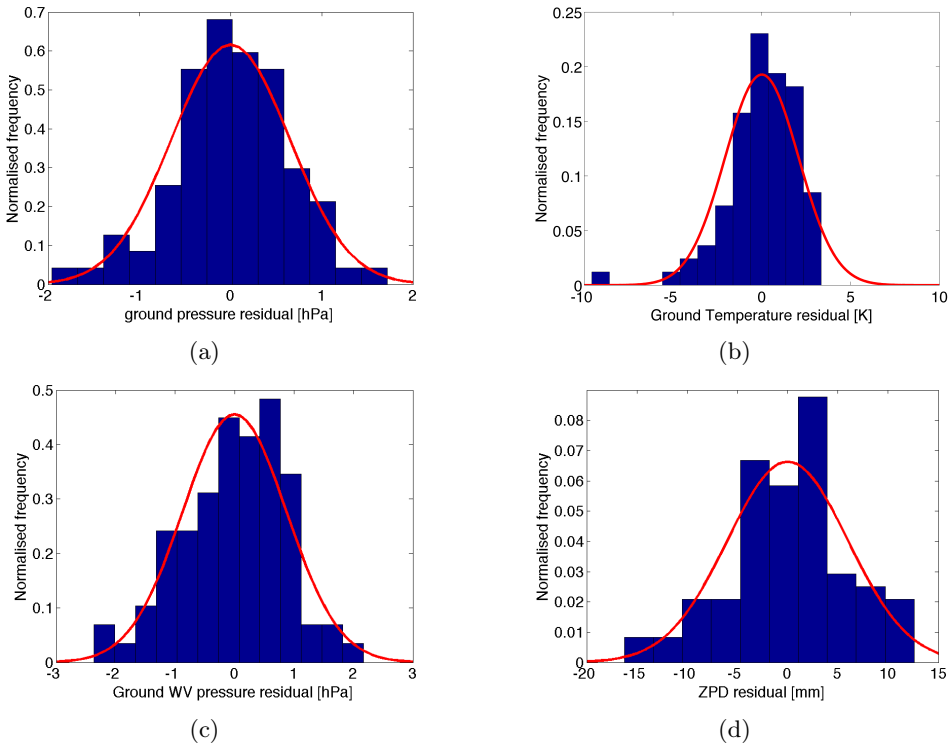


Figure 4.8: Distributions of residual of model quantities minus measured parameters (blue normalised histograms). In red are reported the resulting normal distributions using the variances values. Period is winter at 12:00 UTC. Station is medi (Medicina).

can be computed as the product of the four normal distributions:

$$p(\tilde{P}_0|P_0) = \frac{1}{\sigma_P\sqrt{2\pi}} e^{-\frac{(\alpha_P+\beta_P P_0-\tilde{P}_0)^2}{2\sigma_P^2}}$$

$$p(\tilde{T}_0|\tilde{P}_0 T_1) = \frac{1}{\sigma_T\sqrt{2\pi}} e^{-\frac{(\alpha_T+\beta_T T_1+\gamma_T \tilde{P}_0-\tilde{T}_0)^2}{2\sigma_T^2}}$$

$$p(\tilde{e_{w0}}|\tilde{P}_0\tilde{T}_0 e_{w1}) = \frac{1}{\sigma_e\sqrt{2\pi}} e^{-\frac{(\alpha_e+\beta_e e_{w1}+\gamma_e \tilde{T}_0+\eta_e \tilde{P}_0-\tilde{ew}_0)^2}{2\sigma_e^2}}$$

$$p(\tilde{ZPD}_{GNSS}|\tilde{P}_0\tilde{T}_0\tilde{e}_{w0} ZPD_{MERRA}(\mathbf{x}) h_1 T_1 e_{w1} g_{eff}) = \\ = \frac{1}{\sigma_{ZPD}\sqrt{2\pi}} e^{-\frac{(\mu_{ZPD}-\tilde{ZPD}_{GNSS}+\frac{c_1\tilde{P}_0}{g_{eff}})^2}{2\sigma_{ZPD}^2}} \tag{4.26}$$

being:

$$\mu_{ZPD} = \alpha_{ZPD} + \beta_{ZPD}\left(ZPD_{MERRA}(\mathbf{x}) + c_1\frac{P_1 - P_{42}}{g_{eff}}\right) + \\ +\gamma_{ZPD}\left(\left\langle c_2\frac{e_w}{T}\right\rangle + c_3\left\langle\frac{e_w}{T^2}\right\rangle\right)(z_1 - z_0)$$

Thus for any given set of observations $\tilde{\mathbf{y}}$, the value $\mathbf{w_i} = p(\tilde{\mathbf{y}}|\mathbf{x_i})$ can be computed as the product of the four distributions of Eq. 4.26. It represents the statistical weight for the atmospheric state $\mathbf{x_i}$, whose posterior probability is easily calculated, accordingly to Eq. 4.13 as:

$$Pr(\mathbf{x_i}|\tilde{\mathbf{y}}) = \frac{\mathbf{w_i}}{\sum_{\mathbf{i}=1}^{M}\mathbf{w_i}} \tag{4.27}$$

Once this probability is computed for any state of the prior set, the most likely state as well as the subset of most plausible states, whose cumulative probability just exceeds the desired threshold can be straightforwardly selected. Since the posterior probability for a state $\mathbf{x_i}$ directly transfers to any of its components $\mathbf{x_{ij}}$[7], the distribution for the values of each physical quantity (h, T or e_w) at a given atmospheric level can be also calculated from the weights $\mathbf{w_i}$.

4.3.1 Experimentation on localised data

Two types of tests were carried out. In the first experiment the three sites, namely Cagliari, Ajaccio and Medicine, where a GNSS receiver, MERRA data and balloon soundings were available at the same time for the selected period, where selected. The locations of the GNSS receiver and the radiosonde launch point are not coincident for the selected test sites, but are in the same pixel of the MERRA model. The different localisation actually represents a problem as the ground measurements corresponding to the model output should take into account the different height to which they refers to, in particular as regards the pressure, for which the model provides directly the surface value. However the height of the radiosonde station launch sites and that of the MERRA pixel surface are fixed. Consequently their difference is constant and included in the constant parameter of the linear fit.

Therefore we used the combination of ground level measurements of balloon data and the ZPD value measured from the GNSS receiver as observable measurements:

$$\tilde{\mathbf{y}} = [\tilde{P}_0,\ \tilde{T}_0,\ \tilde{e}_{w0},\ \tilde{ZPD}_{GNSS}] \tag{4.28}$$

In following tables (Table. 4.3, Table. 4.4, Table. 4.5), the resulting parameters of best fit procedure on the three test sites are reported. This set of coefficients must be interpreted as a list of weights that must be used for the linear regression of the variables using Eqs. 4.18 – 4.21. The strongest links of the empirically fitted variables ($\tilde{P}_0$, $\tilde{T}_0$, $\tilde{e}_0$), are exploited from β values very close to the unity.

In most of cases γ_T is not very high, denoting a very low sensitivity of the $\tilde{P}_0$ values from $\tilde{T}_0$, also considering measurements which refers to the same season and time of the day. The opposite for the γ_e coefficient which value shows a known link between $\tilde{e}_{w0}$ and $\tilde{T}_0$. For the troposphere delay parameter modeled using physical relationships is remarkable the fact the the residual quantities given by the quantity $|\alpha_{ZPD}|$ are lower than 10 cm. These results are fully consistent with the working hypotheses. Remarkably the third term on the right hand side of Eq. 4.21, representing the small wet delay of the boundary layer (from ground to the first MERRA level, on average a few hundred of meters in altitude), maintains reasonable values of the order of a few

[7] In general at the generic j-th level some of $\mathbf{x_i}$ state vector can assume the same value. In this case the resulting probability is given by adding the probability assigned to each vector.

Variable	Parameter	Spring		Summer		Autumn		Winter	
		12:00	**00:00**	**12:00**	**00:00**	**12:00**	**00:00**	**12:00**	**00:00**
Pressure [hPa]	α_P	-46.27760	28.20009	30.57028	92.16812	-21.86514	16.02050	-15.37096	19.10060
	β_P	1.06101	0.98698	0.98403	0.92269	1.03686	0.99888	1.03081	0.99623
	σ_P	0.73973	0.96472	0.62411	0.80931	0.78003	0.89668	0.76975	0.81312
Temperature [K]	α_T	102.68331	117.28673	172.15211	101.95639	-57.99792	122.52732	-1.57583	42.08239
	β_T	0.72652	0.92373	0.54581	0.89980	0.93080	1.03084	0.69511	1.31755
	γ_T	-0.02198	-0.09582	-0.03502	-0.07330	0.07776	-0.13262	0.08804	-0.13204
	σ_T	1.39941	1.62618	1.12681	1.21291	1.67335	1.47010	1.03068	2.03669
WV pressure [hPa]	α_e	-322.50588	-198.78723	-347.02289	-259.65807	-216.05947	-188.61449	-129.79160	-49.85776
	β_e	0.77401	0.41078	0.55213	0.36229	0.66875	0.19283	0.68959	0.39512
	γ_e	0.52475	0.31620	0.96493	0.58459	0.45926	0.62685	0.37862	0.31468
	η_e	0.17265	0.11247	0.06988	0.09715	0.08624	0.01878	0.02464	-0.03301
	σ_e	2.13914	1.19846	3.29653	1.97911	1.75648	1.44559	1.32670	0.62783
Trop. Delay [mm]	α_{ZPD}	-12.00147	-14.46234	-1.85561	-6.04267	-17.55366	-9.82405	-8.93223	-8.15045
	β_{ZPD}	0.97755	0.93102	0.88782	0.92671	0.90695	0.95703	0.89535	0.89148
	γ_{ZPD}	0.24177	-0.09530	0.29089	-0.23027	0.76041	-0.34358	0.15541	-0.19071
	σ_{ZPD}	8.96729	8.23957	13.23806	10.95924	8.91654	9.67917	6.55325	6.44864

Table 4.3: Best fit parameters and standard deviations for the computation of the distribution $P(\tilde{y}|xI)$. Site of Ajaccio (ajac). Period of analysis is 2011 full year.

Variable	Parameter	Spring		Summer		Autumn		Winter	
		12:00	**00:00**	**12:00**	**00:00**	**12:00**	**00:00**	**12:00**	**00:00**
Pressure [hPa]	α_P	33.01472	28.47263	6.26501	12.57468	-0.69827	19.58198	0.47363	6.35120
	β_P	0.97162	0.97629	0.99790	0.99181	1.00538	0.98509	1.00425	0.99808
	σ_P	1.14901	0.94767	1.06166	0.63410	0.97420	1.06670	1.27782	0.93461
Temperature [K]	α_T	-79.07990	37.17977	127.29524	153.61618	-93.75407	-30.59865	-32.31861	-9.54828
	β_T	0.92672	1.05306	0.73178	0.82188	1.05943	1.19048	0.94795	1.14825
	γ_T	0.10044	-0.05203	-0.04423	-0.09937	0.07579	-0.02606	0.04690	-0.03400
	σ_T	1.69520	1.47633	1.79865	1.48306	1.51087	1.62031	1.44360	1.76067
WV pressure [hPa]	α_e	-114.07081	-214.11874	-233.34358	-288.67808	-127.61497	-188.32252	-106.07155	-156.56415
	β_e	0.85453	0.35066	0.90802	0.43625	0.83152	0.42956	0.76388	0.29483
	γ_e	0.12273	0.53046	0.13845	0.84825	0.18386	0.51919	0.18959	0.43878
	η_e	0.08090	0.06929	0.19310	0.04896	0.07680	0.04616	0.05474	0.03891
	σ_e	2.70028	2.04231	4.58744	3.11689	2.51534	1.66252	1.50961	1.14886
Trop. delay [mm]	α_{ZPD}	-48.25669	-54.50458	-46.97429	-60.82720	-59.43277	-60.40549	-53.13591	-51.90521
	β_{ZPD}	0.87674	0.89952	0.89648	0.93411	0.91831	0.88982	0.87835	0.89684
	γ_{ZPD}	0.47790	0.27787	0.51982	0.23561	0.66075	0.55160	0.32657	0.01256
	σ_{ZPD}	10.05638	10.49271	16.05577	12.97637	10.95363	11.15340	7.71014	8.16562

Table 4.4: Best fit parameters and standard deviations for the computation of the distribution $P(\tilde{y}|xI)$. Site of Cagliari (cagl). Period of analysis is 2011 full year.

Variable	Parameter	Spring		Summer		Autumn		Winter	
		12:00	**00:00**	**12:00**	**00:00**	**12:00**	**00:00**	**12:00**	**00:00**
Pressure [hPa]	α_P	-160.79886	-15.83160	814.89825	-346.46327	-58.30426	-60.99658	3.41766	15.24579
	β_P	1.18433	1.03967	0.20900	1.36939	1.08235	1.08499	1.02133	1.00948
	σ_P	0.60102	0.42418	0.39858	0.42363	0.65822	0.60692	0.64753	0.58398
Temperature [K]	α_T	-12.29678	171.64493	273.83295	773.75981	-241.90348	124.93583	-7.92170	256.48956
	β_T	0.98831	0.91824	1.24387	0.70803	1.02931	0.83016	0.95659	0.57040
	γ_T	0.01713	-0.14810	-0.33648	-0.67599	0.22750	-0.07769	0.01944	-0.13700
	σ_T	0.70808	1.88188	0.28566	1.99367	1.68969	2.76421	2.06862	2.55868
WV pressure [hPa]	α_e	22.27165	-316.62702	-770.17653	25.65086	-40.08102	-139.17934	-64.56337	-105.29135
	β_e	1.26866	0.23714	0.39598	0.60834	0.88313	0.24276	0.72217	0.22106
	γ_e	-0.08263	0.61260	0.14585	0.33173	0.14553	0.47913	0.18838	0.37320
	η_e	0.00140	0.14831	0.71849	-0.11254	0.00120	0.01197	0.01416	0.00753
	σ_e	0.82214	0.37614	1.37909	0.89135	0.96632	0.67210	0.87609	0.45105
Trop. delay [mm]	α_{ZPD}	-2.89556	7.04142	-19.03134	14.50916	-9.78892	-3.61312	-0.19398	-2.39698
	β_{ZPD}	1.25742	1.01946	0.67709	1.01955	0.92588	0.90336	0.99021	0.95016
	γ_{ZPD}	-0.81918	-0.08417	2.54600	-0.48967	0.86354	0.65446	0.01770	0.21125
	σ_{ZPD}	5.51754	5.82265	5.91781	9.07848	6.95117	7.07642	6.02198	6.68656

Table 4.5: Best fit parameters and standard deviations for the computation of the distribution $P(\tilde{y}|xI)$. Site of Medicina (medi). Period of analysis is 2011 full year.

centimeters after the best fit adjustment. The high seasonal and daily variability of coefficients confirms the necessity to separate between season and between time of the day the dataset. Moreover is remarkable the variability between different MERRA pixel positions, that also indicate a spatial variability. It must be noted that the number of profiles constituting the prior dataset strongly affects the values of fit parameters and the correlations between measures and quantities reconstructed using linear regression, as will be better explained in § 5.2.

In this first test the fitting parameters are only applicable on the site under study and cannot be exported on a wider area or in the MERRA pixels where ground-based measurement of a weather station and a GNSS receiver are available for the period selected for the construction of the prior probability distribution.

4.3.2 Application to the extended area

In the second experiment we tried to explore the applicability of the method over a wider area. IGS stations are userd for GNSS data. For each GNSS station, four MERRA model pixels have been collected (the coincident and the three next neighbor)[8]. In each of these pixels we have also collected data of all the available weather stations inside the pixel area.

Ground-based weather stations measurements are generally at different height, so the height must be included as an additional free parameter in the first three fitting models. Namely through γ_P, η_T, ζ_e (the fourth fitting model already contained the dependence on the height of the weather station z_0). The height differences between the weather station and the MERRA ground surface has been used. Since the ground surface height values of the MERRA pixels are not explicitly given, we simply considered the see surface as lower MERRA level and consequently we replaced the surface pressure (PS) with the sea level pressure (SLP) $z = 0$. Therefore, the state vector has been changed in this second experiment (only in the form) in:

$$\mathbf{x} = [SLP,\ h_1,\ T_1,\ e_{w1},\ h_2,\ T_2,\ e_{w2},\ ...,\ h_{42},\ T_{42},\ e_{w42}] \tag{4.29}$$

In addition, the different height of the receiver sites must be also considered in this second test. In fact in the previous case the values of the coefficients of Tables. 4.3, 4.4 and 4.5 were calculated for each station, the height of the station was a constant implicitly included in the coefficient α_{ZPD}, while in this case all the GNSS receiver measurement are used to construct an unique set coefficient, usable in the extended

[8]The distance is between the GNSS station site and the MERRA pixel centre.

area. Eqs. 4.18 – 4.22 thus become:

$$\tilde{P}_0 = \alpha_P + \beta_P P_0 + \gamma_P \; z_{weather} \tag{4.30}$$

$$\tilde{T}_0 = \alpha_T + \beta_T T_1 + \gamma_T \tilde{P}_0 + \eta_T \; z_{weather} \tag{4.31}$$

$$\tilde{e}_{w0} = \alpha_e + \beta_e e_{w1} + \gamma_e \tilde{T}_0 + \eta_e \tilde{P}_0 + \zeta_e \; z_{weather} \tag{4.32}$$

$$Z\tilde{P}D_{GNSS} - \frac{c_1 \tilde{P}_0}{g_{eff}} = \alpha_{ZPD} + \beta_{ZPD} \left(ZPD_{MERRA}(\mathbf{x}) - c_1 \frac{P_1 - P_{42}}{g_{eff}} \right) + + \gamma_{ZPD} \left(\left\langle c_2 \frac{e_w}{T} \right\rangle + c_3 \left\langle \frac{e_w}{T^2} \right\rangle \right) (z_1 - z_{weather}) + \eta_{ZPD} \; z_{GNSS} \tag{4.33}$$

where $z_{weather}$ and z_{GNSS} are the heights of the weather station and GNSS receiver site respectively.

Also in this case the coefficients have been computed by means of linear regression on the dataset values. The large dataset has allowed to use a high number of measurement-status pairs, as we can see in the scatter plots (see Fig. 4.9).

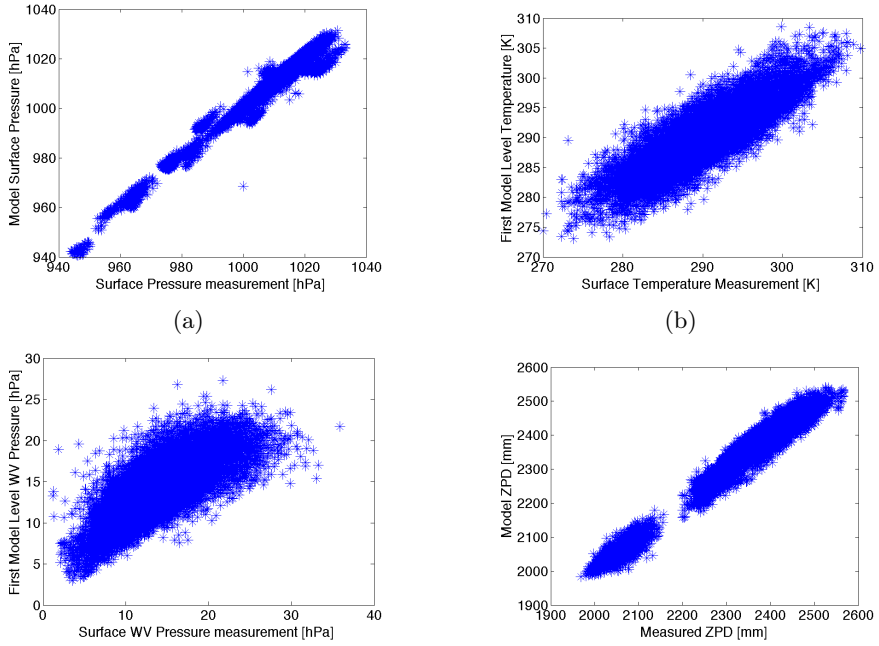


Figure 4.9: Scatter plots of model variables vs. measured parameters for all the area under study (see Fig. 4.1). Period is winter at 12:00 UTC.

The scatter plot of the pressure values presents an apparent anomaly of an overlap of a series of parallel bands (see Fig. 4.9(a)). Actually it is not an anomaly but it is caused by the fact that in the dataset were considered more instruments to measure the pressure on the ground (as opposed to the previous case where a single meteorological station was used), each of which characterised by its accuracy, but also by a different spatial position in the MERRA pixel. Concerning the position, the only additional variable that has been considered is the height of the weather station, which for example, directly affects the scatter plot of the pressure. In the specific case by

not considering the dependence on height, the dispersion of the pressure scatter plot increases a lot exhibiting more separated groups. A similar consideration can be made for the ZPD, as different GNSS stations at different heights give measurements of ZPD which systematically differ due to the different heights. The measured and model ZPD values show a very strong correlation, but also a gap in the scatter plot (see Fig. 4.9(d)), which is precisely due to the different height of the stations (in particular the IGS station named gras sited at an altitude of 1319 m is the cause of low values of ZPD, which are spaced from the other).

The scatter plots for all the area show a greater variability, mainly as regards temperature and partial pressure of water vapour.

As for the individual stations, also in this case we have analysed the distributions of the residuals (difference between the measured value and the value reconstructed by the model). The normalised histograms of Fig. 4.10 represents this distribution, that also in this case appear well represented and approximated by Gaussian distributions:

$$p(\tilde{P}_0|P_0\ z_{weather}) = \frac{1}{\sigma_P\sqrt{2\pi}} e^{-\frac{(\alpha_P+\beta_P P_0+\gamma_P\ z_{weather}-\tilde{P}_0)^2}{2\sigma_P^2}}$$

$$p(\tilde{T}_0|\tilde{P}_0 T_1\ z_{weather}) = \frac{1}{\sigma_T\sqrt{2\pi}} e^{-\frac{(\alpha_T+\beta_T T_1+\gamma_T\tilde{P}_0+\eta_T\ z_{weather}-\tilde{T}_0)^2}{2\sigma_T^2}}$$

$$p(\tilde{e_{w0}}|\tilde{P}_0\tilde{T}_0 e_{w1}\ z_{weather}) = \frac{1}{\sigma_e\sqrt{2\pi}} e^{-\frac{(\alpha_e+\beta_e e_{w1}+\gamma_e\tilde{T}_0+\eta_e\tilde{P}_0+\zeta_e\ z_{weather}-\tilde{e_{w0}})^2}{2\sigma_e^2}}$$

$$p(\tilde{ZPD}_{GNSS}|\tilde{P}_0\tilde{T}_0\tilde{e}_{w0} ZPD_{MERRA}(\mathbf{x}) h_1 T_1 e_{w1} g_{eff}\ z_{weather}\ z_{GNSS}) = \\ = \frac{1}{\sigma_{ZPD}\sqrt{2\pi}} e^{-\frac{(\mu_{ZPD}-\tilde{ZPD}_{GNSS}+\frac{c_1\tilde{P}_0}{g_{eff}})^2}{2\sigma_{ZPD}^2}} \tag{4.34}$$

being:

$$\mu_{ZPD} = \alpha_{ZPD} + \beta_{ZPD}\left(ZPD_{MERRA}(\mathbf{x}) + c_1\frac{P_1 - P_{42}}{g_{eff}}\right) + \\ +\gamma_{ZPD}\left(\left\langle c_2\frac{e_w}{T}\right\rangle + c_3\left\langle\frac{e_w}{T^2}\right\rangle\right)(z_1 - z_{weather}) + \eta_{ZPD}\ z_{GNSS}$$

The likelihood probability can be computed starting from the product of this four distributions given a set of observables $\tilde{\mathbf{y}}$ and for each status $\mathbf{x_i}$, retrieving the quantity $\mathbf{w_i} = p(\tilde{\mathbf{y}}|\mathbf{x_i})$ and applying Eq. 4.27.

The availability of a more statistically significant sample allows to build a set of coefficients that is more representative of the study area as a whole, and consequently allows to circumvent the problem of the fluctuations of the coefficients when moving from one point to another of the map. The consequent set of coefficients is listed in Table 4.6 that shows the stability of coefficients when passing from one season to the other but also when different time of the day is considered.

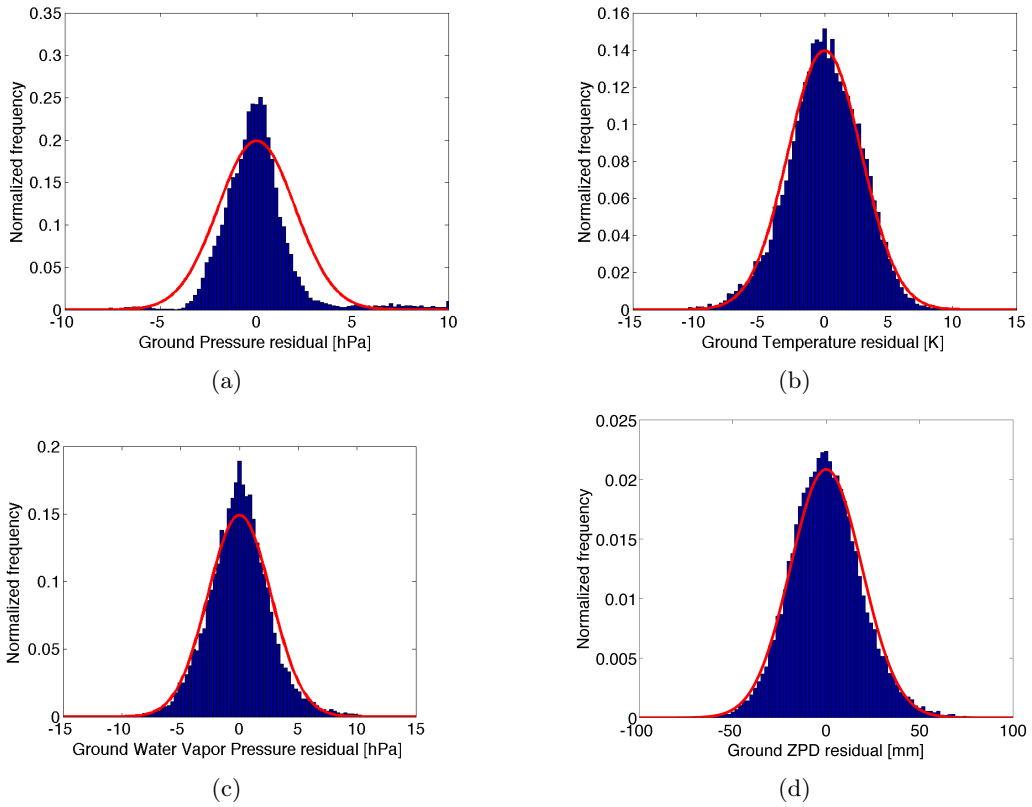


Figure 4.10: Distributions of residual of model quantities minus measured parameters (blue normalised histograms), valid for the whole area under study. In red are reported the resulting normal distributions using the variances values. Period is winter at 12:00 UTC.

The main disadvantage is the increment of the variances for all the modeled parameters, which includes also different instruments error variances beyond the variability of parameters relative to different areas. The effects of this increment of the variance is clearly evident by analyzing the coefficient values. However when considering the pressure model, as shown in the residual probability distributions (see Fig. 4.10(a)), the nominal normal distribution could be effectively more narrow and tall. The enlargement of the distribution is due to a series of residual values between 5 and 10 and below -5 hPa, which are the cause of a greater variance, while having an almost negligible probability.

As we have made a test of applicability over areas not covered by measurements, we have excluded from the dataset the observations relative to the test site, in the step of likelihood distribution reconstruction. This is a test of the robustness of the method, which allows to evaluate the effectiveness of the retrieval using also states $\mathbf{x_i}$ not included in the dataset used for the fitting, finalised to the calculation of the coefficients. In other words in the process of calculation of the weight functions $\mathbf{w_i}$, the observable measurement and the atmospheric states relative to the site on which the retrieval has been applied have been excluded from the set of $\tilde{\mathbf{y}}$ and $\mathbf{x_i}$ respectively. This dataset have been instead used in the process of retrieval.

Variable	Parameter	Spring		Summer		Autumn		Winter	
		12:00	**00:00**	**12:00**	**00:00**	**12:00**	**00:00**	**12:00**	**00:00**
Pressure [hPa]	α_P	-72.42800	-66.59250	-66.23609	-68.60787	-62.43207	-53.70745	-45.33258	-38.84155
	β_P	1.07161	1.06564	1.06529	1.06759	1.06151	1.05291	1.04467	1.03828
	γ_P	-0.12133	-0.12022	-0.11988	-0.11921	-0.11913	-0.11930	-0.11967	-0.12018
	σ_P	2.08373	2.00443	1.91281	1.84993	1.92020	2.05241	2.14119	2.20091
Temperature [K]	α_T	64.72610	-18.75959	12.99033	-10.14306	-3.64524	-28.99471	-28.47521	-33.22600
	β_T	0.87898	1.05564	0.93530	1.03228	1.05025	1.04853	1.06467	1.04918
	γ_T	-0.03092	0.00284	0.00567	0.00099	-0.01081	0.01469	0.00942	0.01858
	η_T	-0.00574	-0.00381	-0.00322	-0.00437	-0.00603	-0.00336	-0.00413	-0.00295
	σ_T	2.40096	2.85545	2.75452	2.82664	2.82669	2.88041	2.86711	2.93749
WV pressure [hPa]	α_e	-139.21650	-96.96186	-101.02928	-107.62159	-105.60362	-103.96712	-107.59125	-100.67295
	β_e	0.37499	0.66063	0.68907	0.68800	0.68534	0.67094	0.66130	0.66461
	γ_e	0.50192	0.26822	0.29121	0.29689	0.31773	0.32268	0.32108	0.31251
	η_e	0.00389	0.02471	0.02198	0.02686	0.01894	0.01604	0.02023	0.01581
	ζ_e	-0.00330	-0.00050	-0.00114	-0.00001	-0.00061	-0.00058	0.00020	-0.00015
	σ_e	2.10614	2.67693	2.93036	3.44242	3.43616	3.20002	3.01189	2.83427
Trop. Delay [mm]	α_{ZPD}	46.98862	48.55926	47.30242	46.65926	46.78254	47.38704	47.47183	46.97506
	β_{ZPD}	0.87127	0.88419	0.88986	0.90327	0.90890	0.91486	0.92135	0.93135
	γ_{ZPD}	-3.78414	-3.78026	-3.45739	-3.14955	-3.18949	-3.35883	-3.52143	-3.74927
	η_{ZPD}	-0.28537	-0.28492	-0.28549	-0.28716	-0.28711	-0.28738	-0.28664	-0.28531
	σ_{ZPD}	18.49738	19.13456	19.78710	20.71566	21.21176	21.47743	21.84811	22.31592

Table 4.6: Best fit parameters and standard deviations for the computation of the distribution $P(\tilde{y}|xI)$ valid for the whole area under study. Period of analysis is full year 2011.

4.3.3 Retrieval of atmospheric profiles

Once the likelihood probability distribution is defined, the retrieval process can be easily applied to all the measurements occurrences $\tilde{\mathbf{y}}$ (at synoptic times, 00, 06, 12, 18 UTC). The possible atmospheric state can be selected from the whole dataset using spatial and temporal criteria. Different spatial and temporal criteria have been applied depending on the type of test. In the first test data relative to MERRA pixel coincident with the weather and the GNSS stations have been collected, of given season and time of the day. In the second test the possible atmospheric states dataset has been constructed by considering data of the 4 most neighbor MERRA pixels of the GNSS stations still of given season and time. The so selected dataset constitutes the prior dataset (in which all the atmospheric states are equiprobable). For each of them the probability is assigned as a weight using Eq. 4.27. The retrieval process consists of ordering the states ($\mathbf{x_i}$) in descending order of their weight and selecting the profiles that allow to get an overall probability (as sum of the weights of the individual profiles) that exceeds a predetermined threshold. This threshold has been set to 0.7 for all the results shown in the present work[9]. In this way the most probable profile (i.e. the one with higher weight) can be identified. In general, it is not necessarily the one closest to reality. Moreover, the weight of the most probable profile could be very close to that of other profiles, then it is more reasonable the choice of a probability threshold. However in some cases the combination of the probability according to the formula of Eq. 4.17 is highly selective, and only a few profiles, if not one, are extracted.

In the experiment the chosen dataset was the 2011 year, whose states were excluded from the 2001-2010 period used as prior ensemble as well as for the empirical parameters estimation. This is an even more stringent condition with respect to what could be a real situation, when reanalysis studies are available up to few days before the current date.

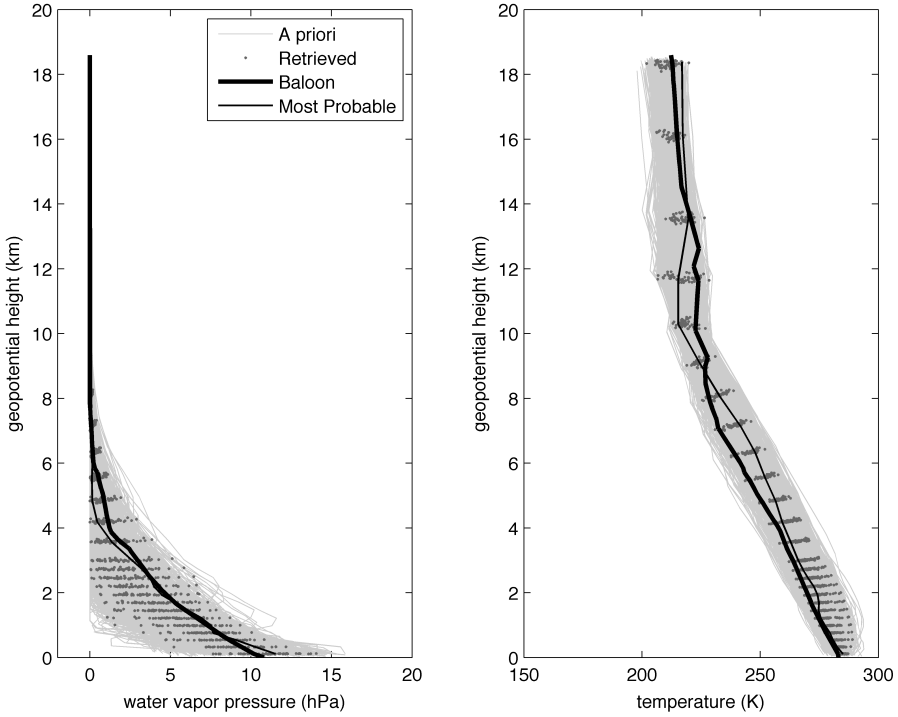
In the results of retrieval of profiles of water vapour and temperature (see next figures) we reported both the most probable profile and the profiles selected by the algorithm limiting the achievement of the cumulative probability to the 0.7 threshold. For the chosen case studies, the results of both the first test (Figs. 4.11 – 4.15) and the second one (Figs. 4.16 – 4.20) are reported. The so selected most probable profiles of temperature and WV are plotted, together with all the prior profiles and the balloon measured values to be used as term of comparison. Finally in order to have an assessment of the added value introduced by the GNSS receiver measurement, two different results are depicted: the first (top figures) is the result of the retrieval when

[9]The value of 0.7 indicate a confidence interval of about $\pm$ 1 σ in normal distributions.

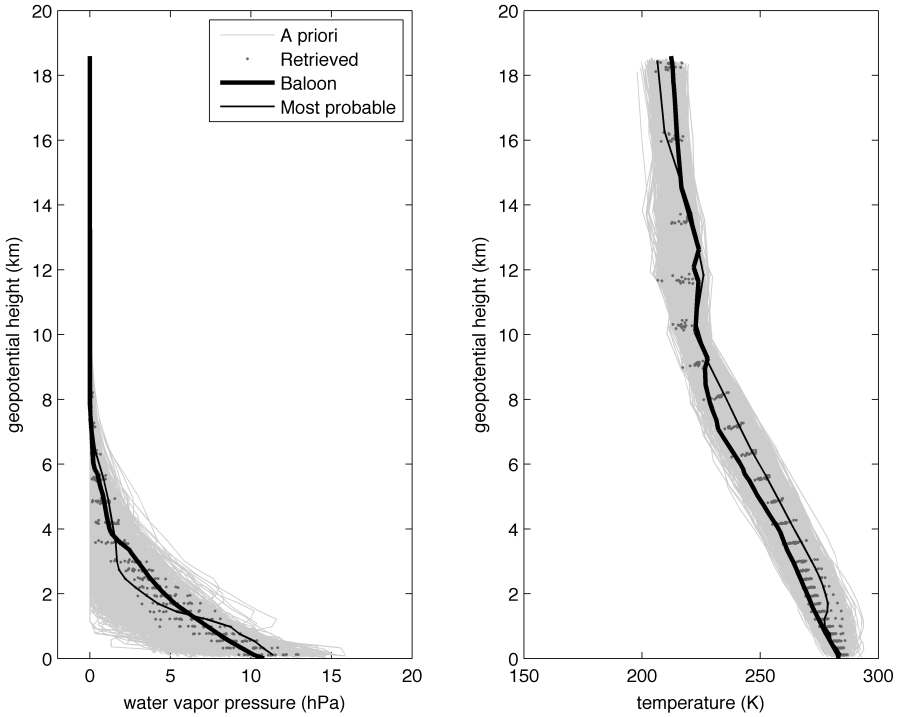
the likelihood is based only on the weather station surface measurements and considering uniform the probability $Pr(Z\tilde{P}D_{GNSS}|\tilde{P}_0\tilde{T}_0\tilde{e}_{w0}\mathbf{x_i})$, the second, (bottom figures), is the result of the retrieval using all the available measurements, including the GNSS receiver ZPD measurements. The most probable retrieved profile is also indicated.

This graphical representation allows a rapid evaluation "by eye". A result that can be considered "poor" is the case shown in Fig. 4.11 and Fig. 4.16. As apparent the retrieved profiles summing up the 70% of cumulative probability nearly span most of the region covered by the whole set of possible profiles, drawn in light grey. Although the most probable retrieved profiles of both WV and temperature match enough those measured by balloon, a great uncertainty affects the retrieval, since a lot of different profiles have almost the same probability. In this case the addition of the GNSS measurement slightly reduces the dispersion of the retrieved states, even if produces worst most probable profiles for both WV and temperature. A result that can be considered "good" is instead shown in Fig. 4.12 and Fig. 4.17. In this case a few profiles sum up the 70% of probability and when the GNSS measurement is included a single WV profile exceeds the probability of 50% alone, as particularly evident in Fig. 4.12. This effect of high selectivity introducing the GNSS observation, although less evident, is also present in the case of the second test (Fig. 4.17). Moreover the most probable profile selected in the two different methods are exactly the same. Indeed this profile seems very close to that measured by the balloon. Good results are also obtained for the temperature profile, which closely follows the measured one. As apparent the agreement is maintained up to high altitudes, more or less corresponding to the upper tropospheric limit ($\simeq$12 km). Despite the information used comes from ground measurements, such an agreement extended to the upper levels was observed in a large number of the examined cases, especially when the GNSS delay was also used.

In some cases as that of Fig. 4.13 the number of selected profiles is already low when only surface measurement are considered. The introduction of the GNSS measurement still narrows the range of variability and selects a profile closer to the balloon measurement, regarding both the WV and the temperature. For the same case the configuration of the second test (Fig. 4.18) provides results that are slightly less selective, due to the variance increment of the linear model fit and the consequent enlargement of the Gaussian curves. However in the WV profiles the introduction of the ZPD measure eliminates profiles very distant from the real one, which are characterised by a high WV content. The most probable profile of WV is quite different from the measured one, indicating a poor retrieval performance. The opposite for the temperature whose retrieved values well follow the vertical distribution of measured values, including the reconstruction of two very little inversions above the surface level. An analysis of the weather situation has been made for to this case study, for trying to understand the

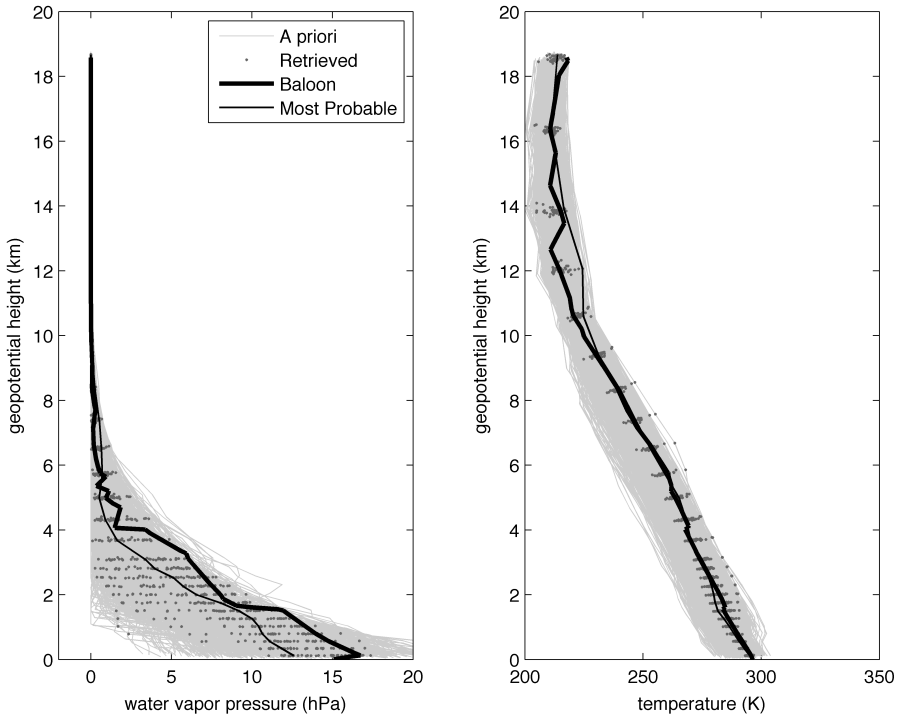


(a) Retrieval result using only surface measurements ($\tilde{P}_0$, $\tilde{T}_0$, $\tilde{e_{w0}}$)

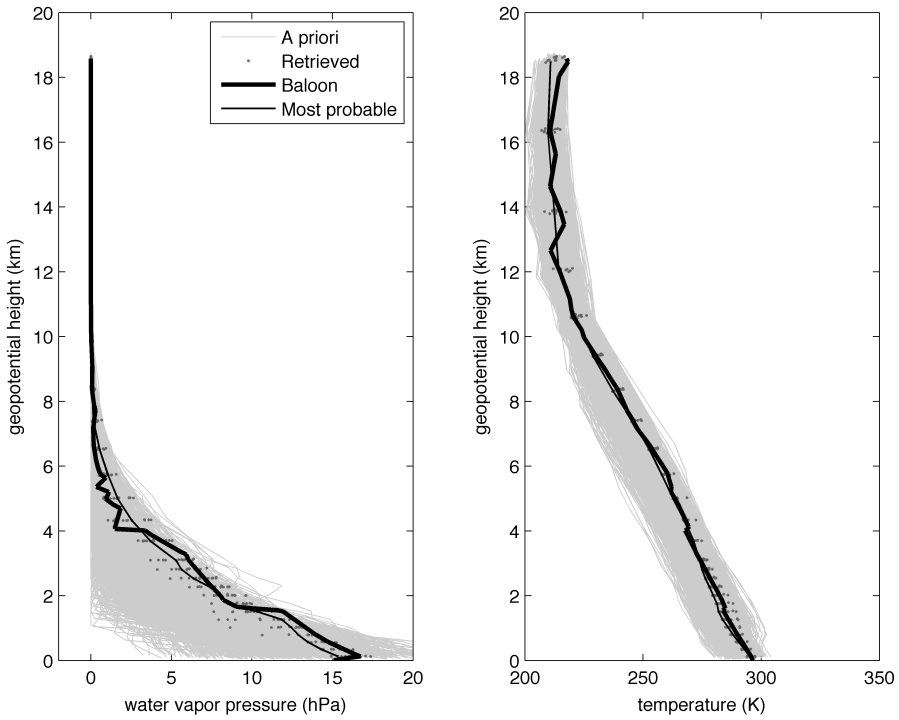


(b) Retrieval result using all measurements ($\tilde{P}_0$, $\tilde{T}_0$, $\tilde{e_{w0}}$, $\tilde{ZPD}$)

Figure 4.11: Retrieved profiles of water vapour pressure, and temperature, using only surface measurements (a) and including the GNSS measurements (b). Retrieval using locally valid coefficients (test one). Date: 2011/02/01 time: 00:00 UTC. Station is cagl (Cagliari).

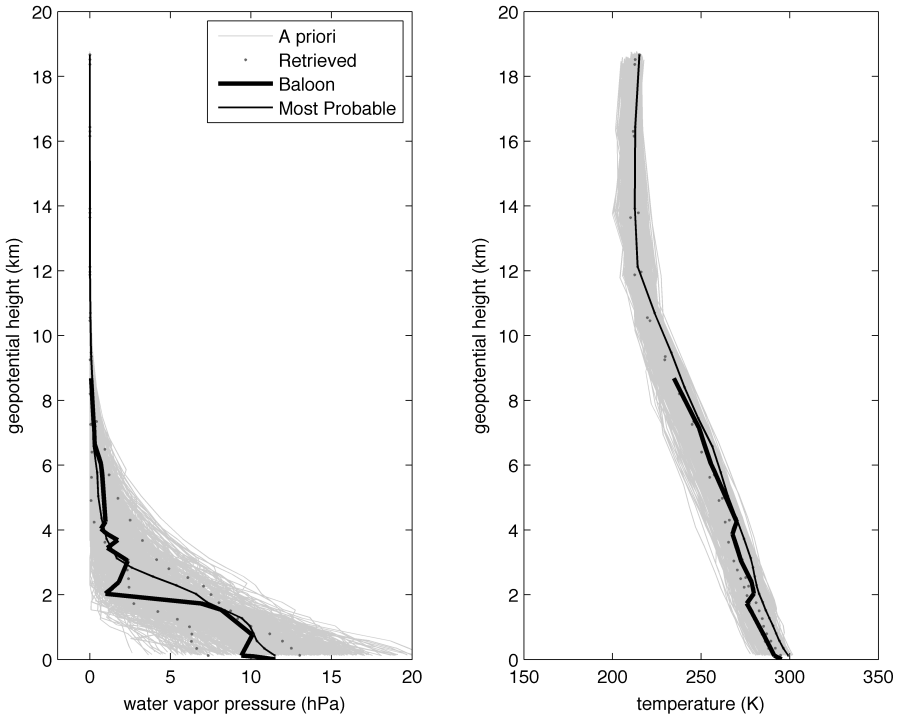


(a) Retrieval result using only surface measurements ($\tilde{P_0}$, $\tilde{T_0}$, $e_{\tilde{w}0}$)

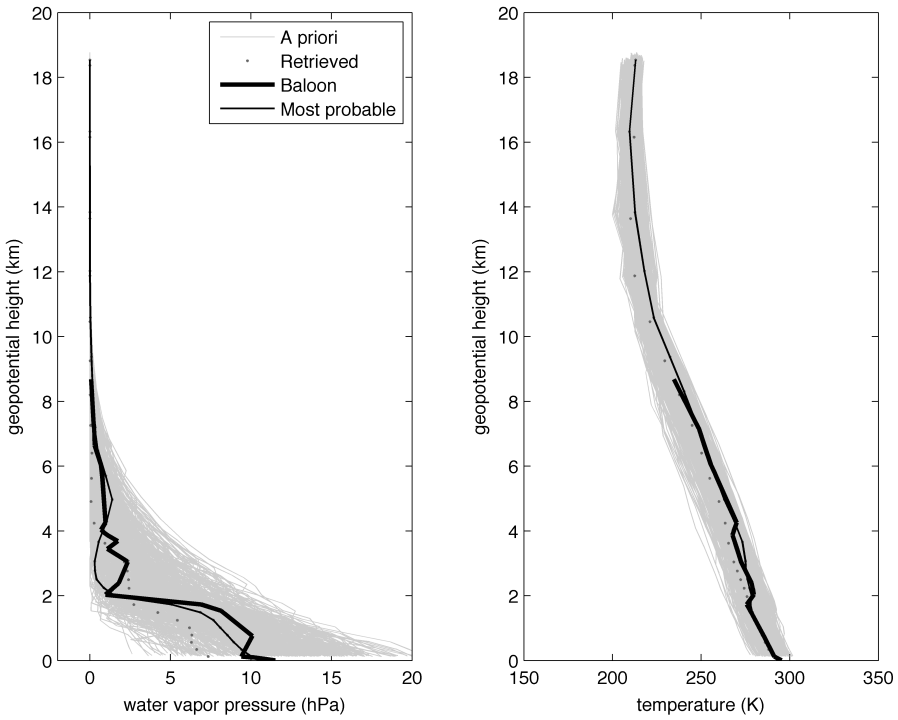


(b) Retrieval result using all measurements ($\tilde{P_0}$, $\tilde{T_0}$, $e_{\tilde{w}0}$, $Z\tilde{P}D$)

Figure 4.12: Same as Fig. 4.11 but date is 2011/10/07; time: 12:00 UTC and station is cagl (Cagliari).

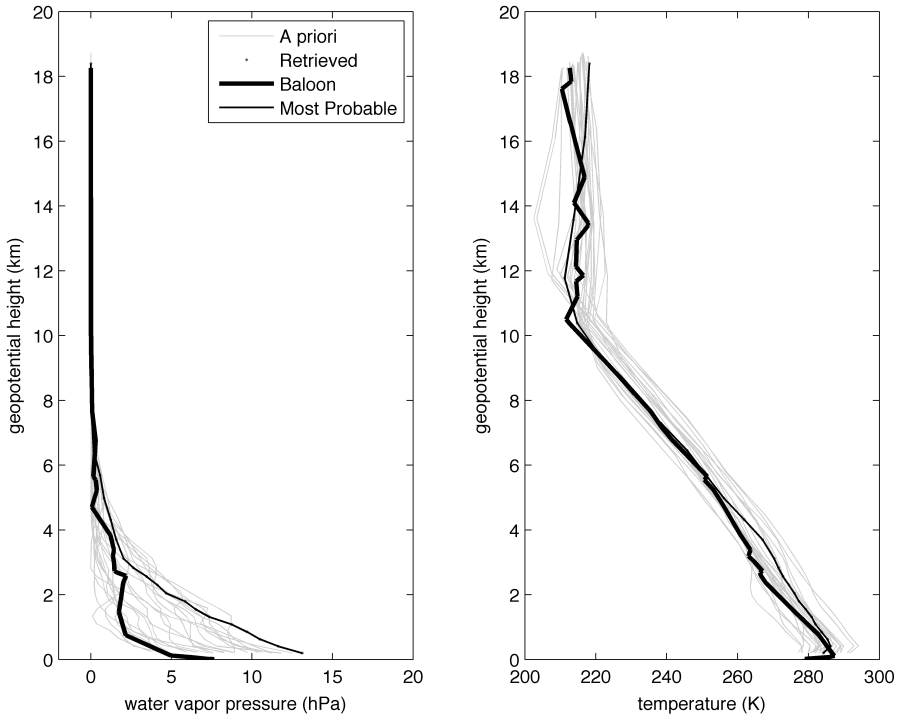


(a) Retrieval result using only surface measurements ($\tilde{P_0}$, $\tilde{T_0}$, $e_{\tilde{w0}}$)

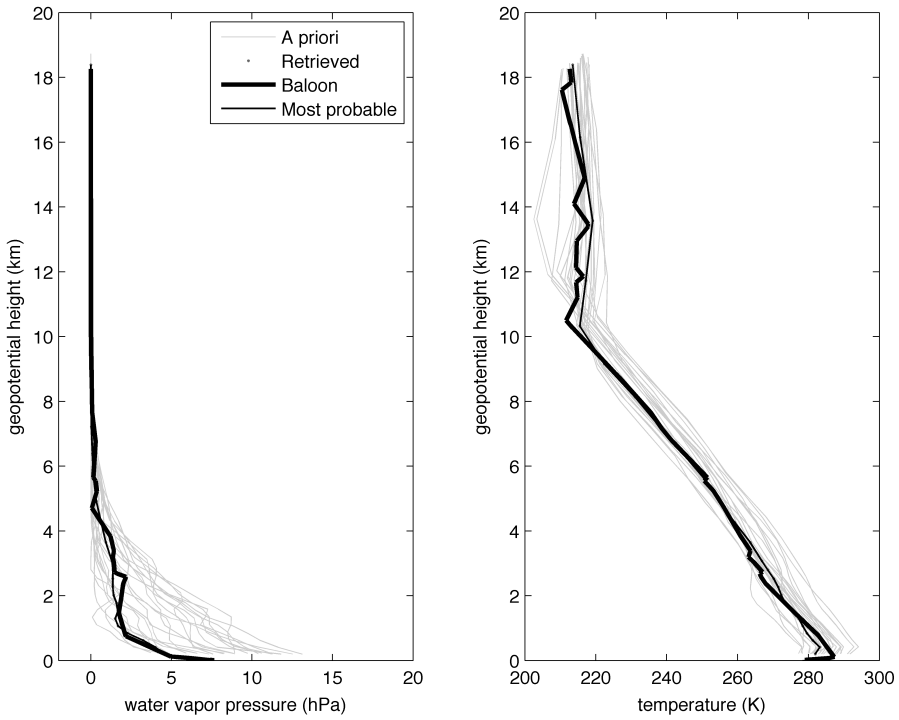


(b) Retrieval result using all measurements ($\tilde{P_0}$, $\tilde{T_0}$, $e_{\tilde{w0}}$, $Z\tilde{P}D$)

Figure 4.13: Same as Fig. 4.11 but date is 2011/10/07; time: 12:00 UTC and tation is ajac (Ajacc).

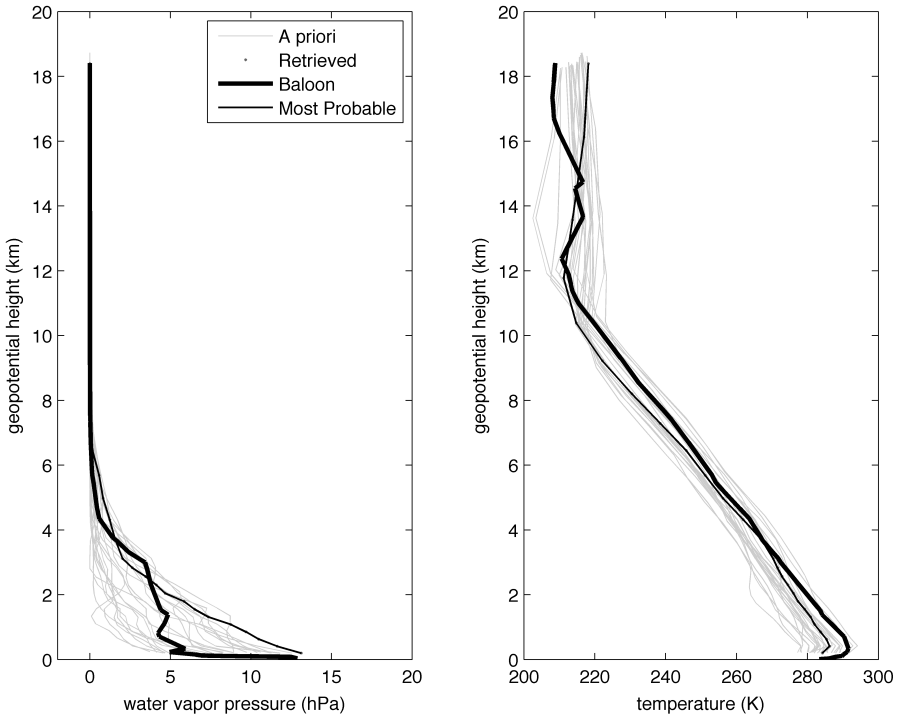


(a) Retrieval result using only surface measurements ($\tilde{P}_0$, $\tilde{T}_0$, $e_{\tilde{w}0}$)

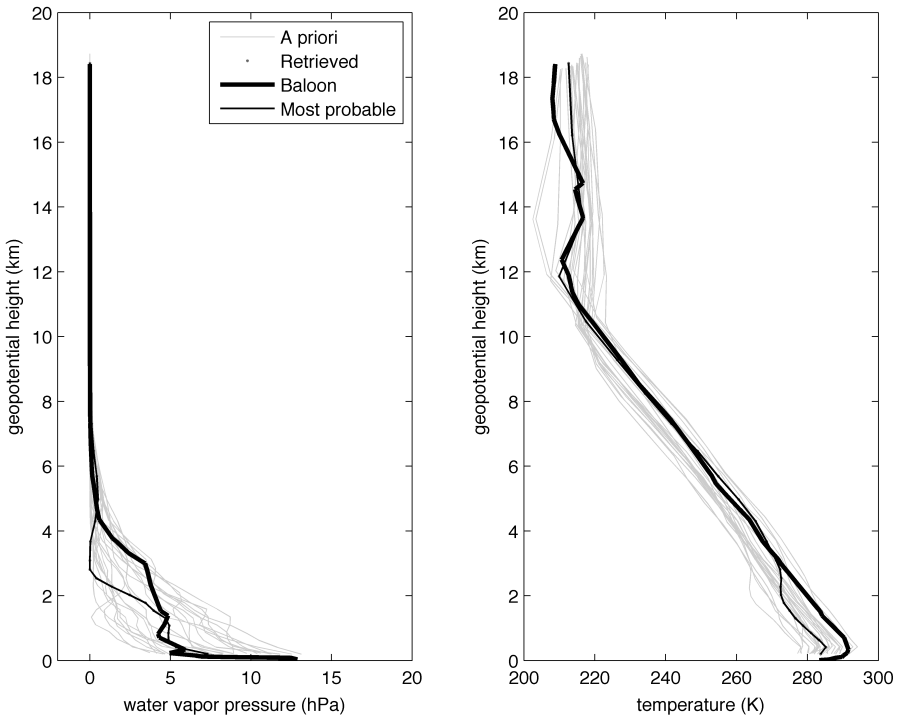


(b) Retrieval result using all measurements ($\tilde{P}_0$, $\tilde{T}_0$, $e_{\tilde{w}0}$, $Z\tilde{P}D$)

Figure 4.14: Same as Fig. 4.11 but date is 2011/04/14; time: 00:00 UTC and station is medi (Medicina).



(a) Retrieval result using only surface measurements ($\tilde{P_0}$, $\tilde{T_0}$, $\tilde{e_{w0}}$)



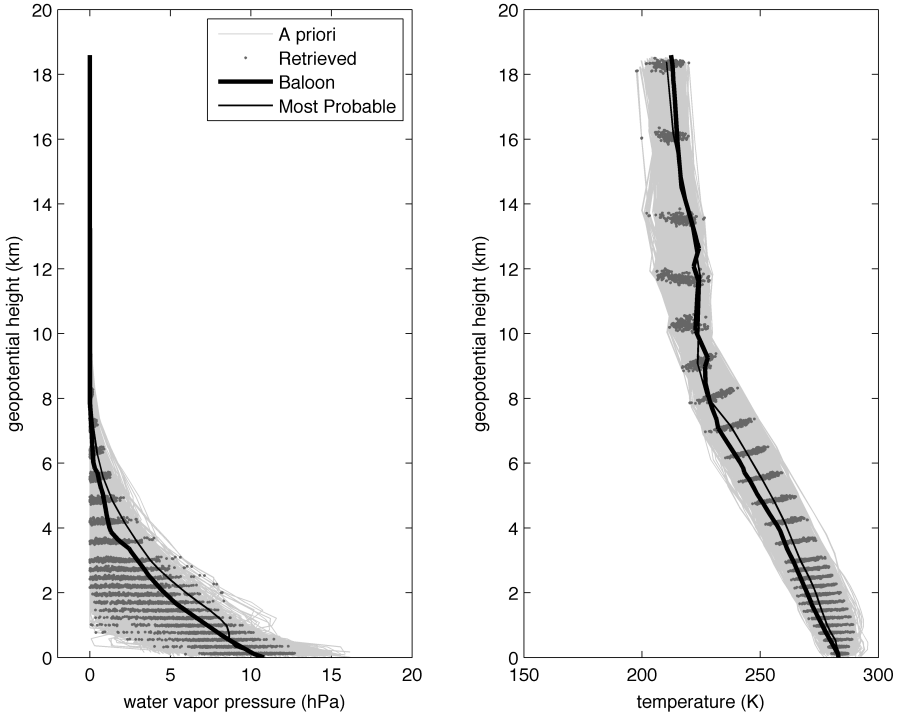
(b) Retrieval result using all measurements ($\tilde{P_0}$, $\tilde{T_0}$, $\tilde{e_{w0}}$, $\tilde{ZPD}$)

Figure 4.15: Same as Fig. 4.11 but date is 2011/04/11; time: 00:00 UTC and station is medi (Medicina).

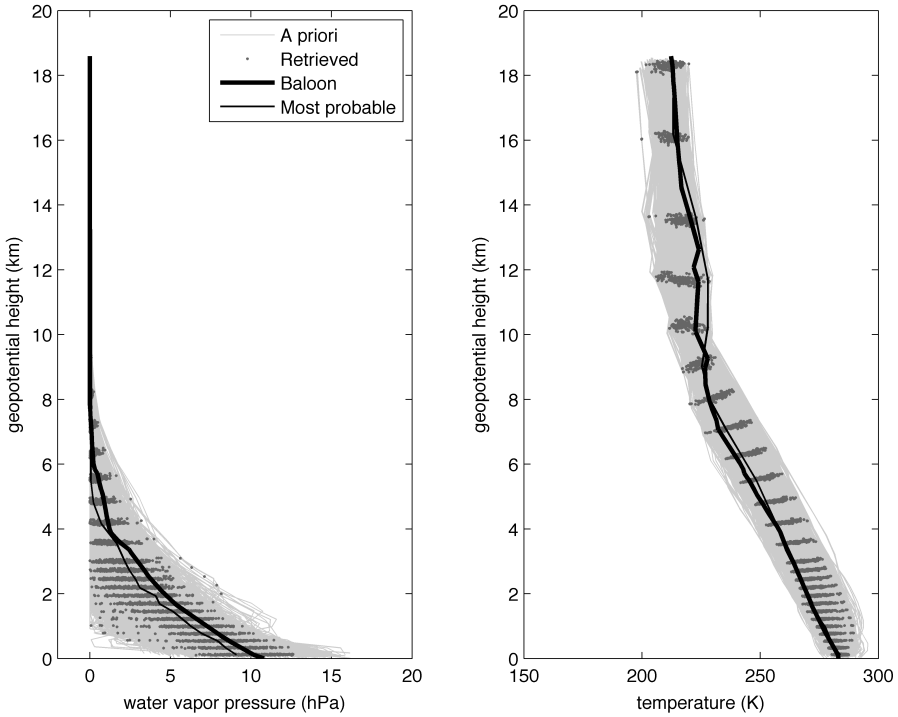
causes that have led to an incorrect reconstruction of the water vapour profile. The synoptic situation shows the presence of a westerly air flow. As the Ajaccio site is located in the west side of the Corsica island, it is subject to this flow of air which as coming from the sea has an high humidity content. The Corsica orography presents a series of mountain quite high, crossing longitudinally from north to south. That orographic pattern interacts with the air flow, by forcing the air to move up and to cool down. At the height of the top mountain level (about 2 km) there is a sort of temperature discontinuity (more properly a strong horizontal gradient), between the western and the eastern sides. That discontinuity is probably seen by the balloon sounding, which is transported by the flow taking measurement relative to the humid western air mass at the lower levels and the more dry eastern air mass at the upper levels. Moreover also the ZPD measurement is affected by this discontinuity and the resulting values underestimates the real water vapour content. It can be hypothesised for example that the horizontal inhomogeneity of the atmosphere and a low number of on view satellites could be the cause of the ZPD underestimation and the distance of from the GNSS profiles.

In the case of retrieval over the Medicina station (Fig. 4.14, Fig. 4.15), the number of available profiles for the prior dataset is very low and insufficient to allow to get satisfactory results in the retrieval process. In fact one of the critical aspects of this type of algorithm is obviously the number of profiles that constitute the prior dataset (see § 5.2), in other words the completeness of the prior.

Nonetheless when the first configuration of the retrieval is applied, actually there are some cases in which the retrieval succeeds quite well, certainly in consequence of the fact that the coefficients of the probability distribution functions are calculated on the basis of the data of the prior dataset. The scarcity of the number of profiles allows to select only one profile in most cases, which reaches a probability of more than 70%. When considering the coefficients with validity over the whole area (Fig. 4.19, Fig. 4.20) the results of application and the effectiveness of the method in this configuration are even more evident. The introduction of the ZPD measurement allows to improve the accuracy of the WV profile, also tightening the uncertainty range with a reduction of the number of selected profiles. The most significative result is that the retrieval is possible where the first test, although more accurate, was not applicable.

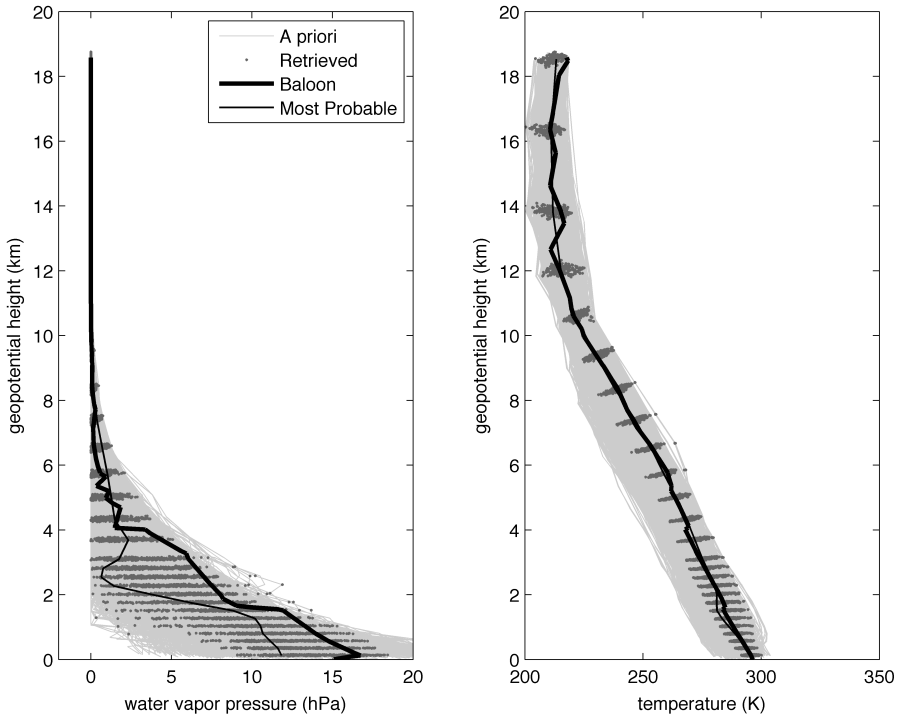


(a) Retrieval result using only surface measurements ($\tilde{P}_0$, $\tilde{T}_0$, $\tilde{e_{w0}}$)

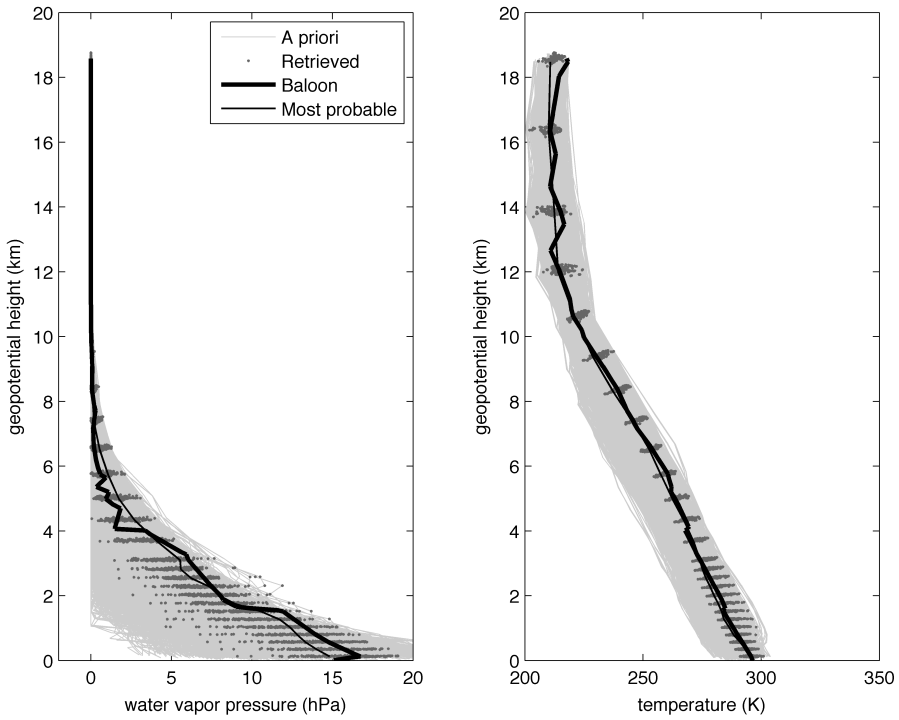


(b) Retrieval result using all measurements ($\tilde{P}_0$, $\tilde{T}_0$, $\tilde{e_{w0}}$, $\tilde{ZPD}$)

Figure 4.16: Retrieved profiles of water vapour pressure, and temperature, using only surface measurements (a) and including the GNSS measurements (b). Retrieval using whole area valid coefficients (test two). Date: 2011/02/01 time: 00:00 UTC. Station is cagl (Cagliari).

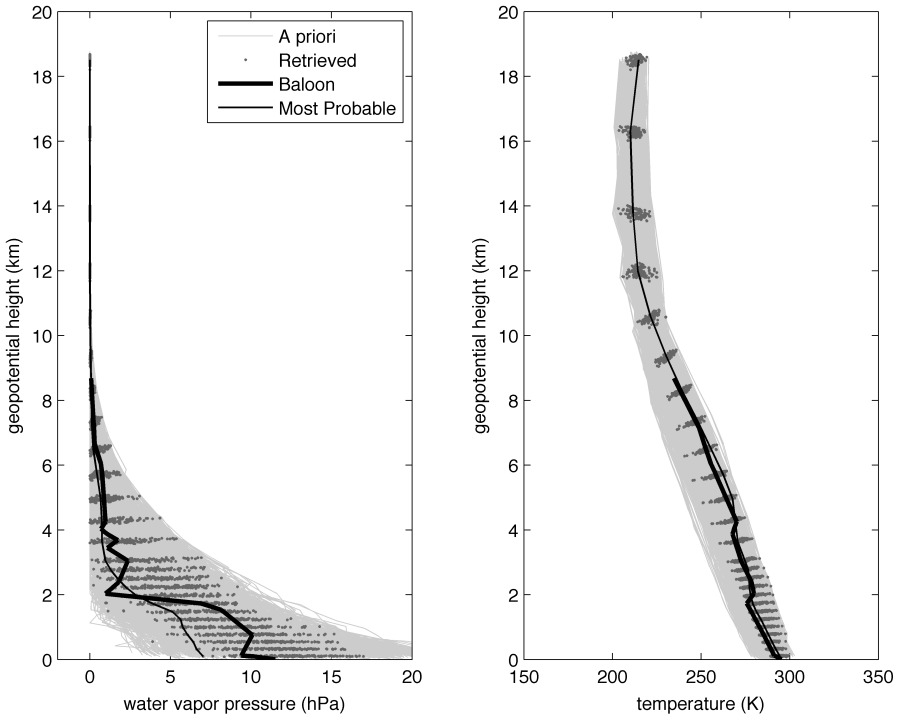


(a) Retrieval result using only surface measurements ($\tilde{P_0}$, $\tilde{T_0}$, $e_{\tilde{w}0}$)

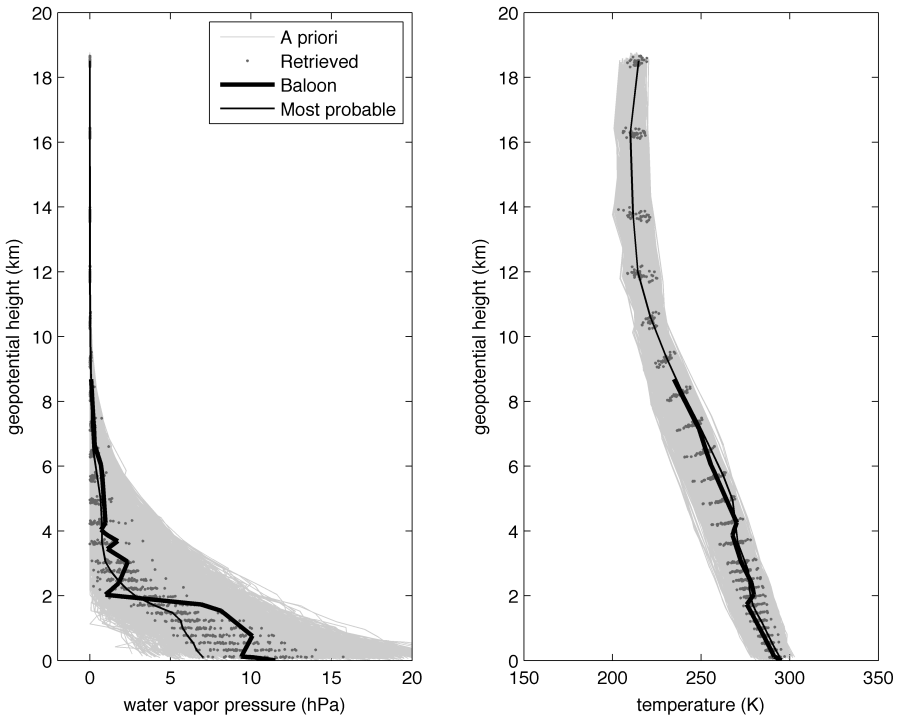


(b) Retrieval result using all measurements ($\tilde{P_0}$, $\tilde{T_0}$, $e_{\tilde{w}0}$, $Z\tilde{P}D$)

Figure 4.17: Same as Fig. 4.16 but date is 2011/10/07; time: 12:00 UTC and station is cagl (Cagliari).

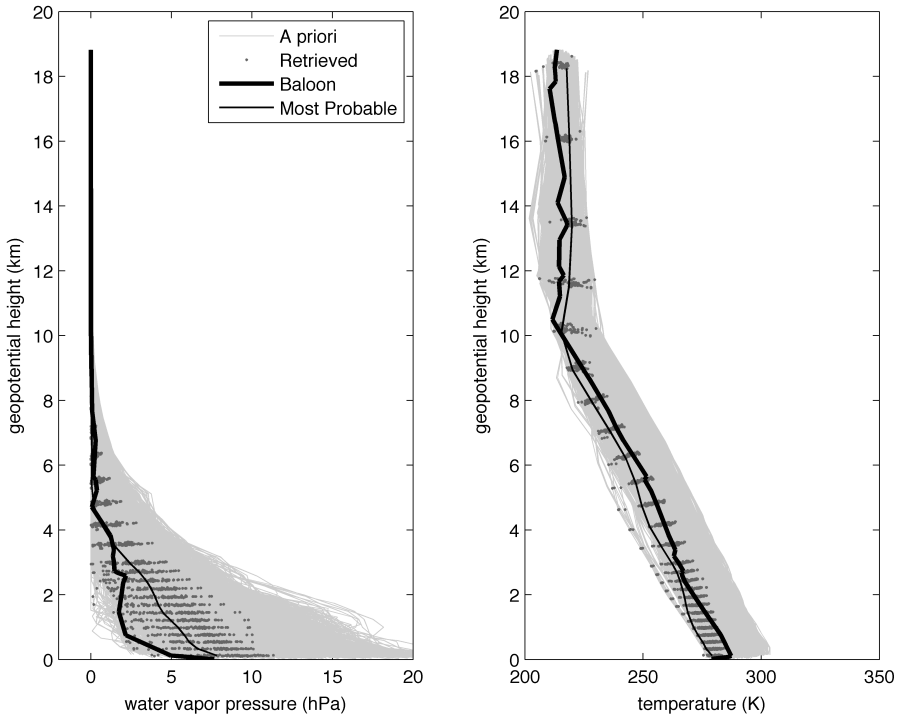


(a) Retrieval result using only surface measurements ($\tilde{P}_0$, $\tilde{T}_0$, $e_{\tilde{w}0}$)

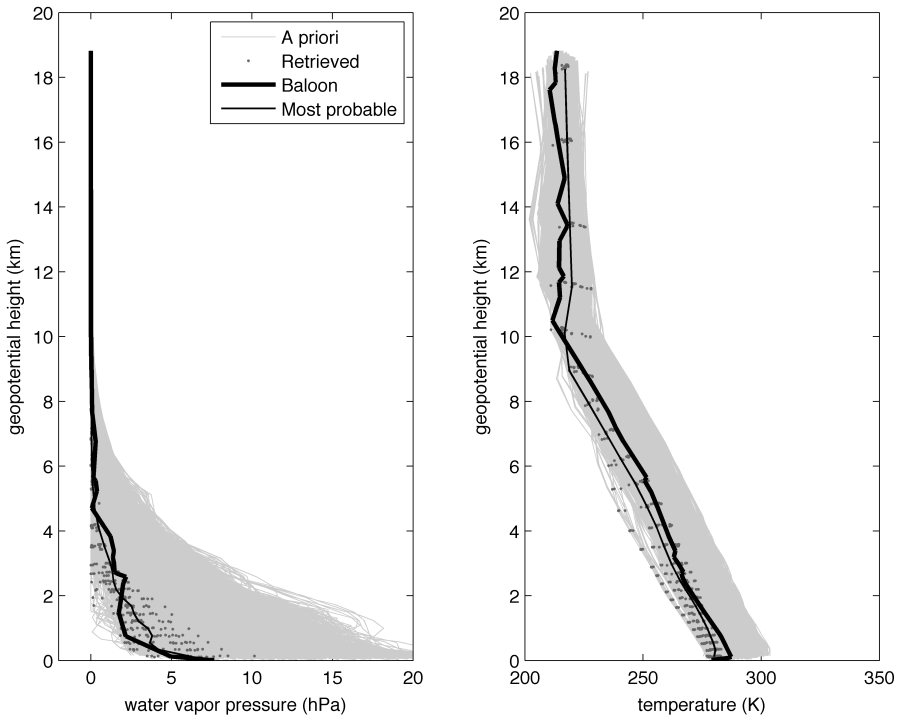


(b) Retrieval result using all measurements ($\tilde{P}_0$, $\tilde{T}_0$, $e_{\tilde{w}0}$, $Z\tilde{P}D$)

Figure 4.18: Same as Fig. 4.16 but date is 2011/10/07; time: 12:00 UTC and station is ajac (Ajacc).

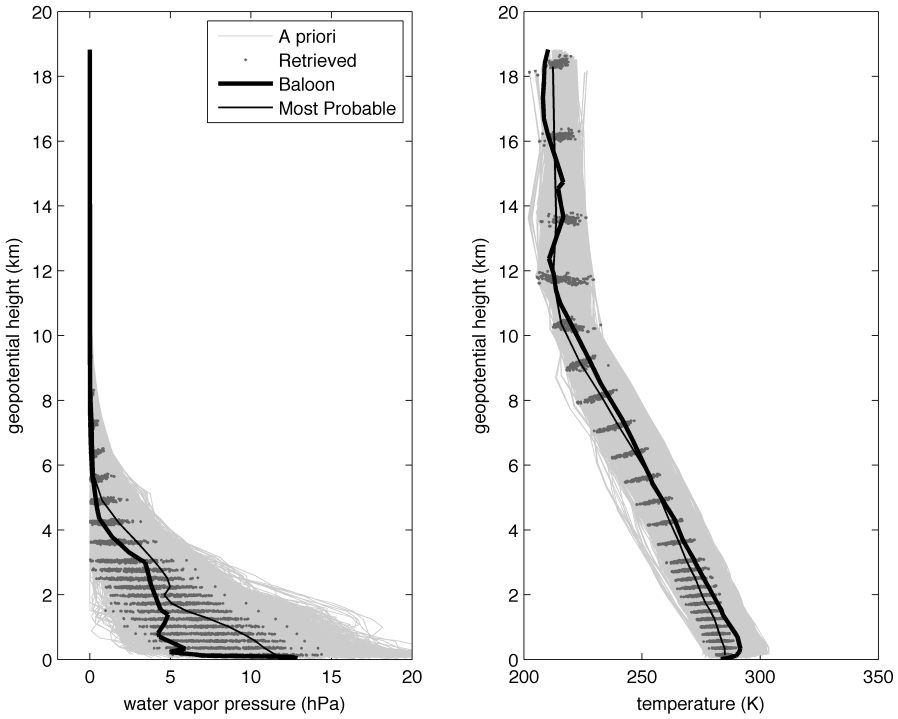


(a) Retrieval result using only surface measurements ($\tilde{P}_0$, $\tilde{T}_0$, $e_{\tilde{w}0}$)

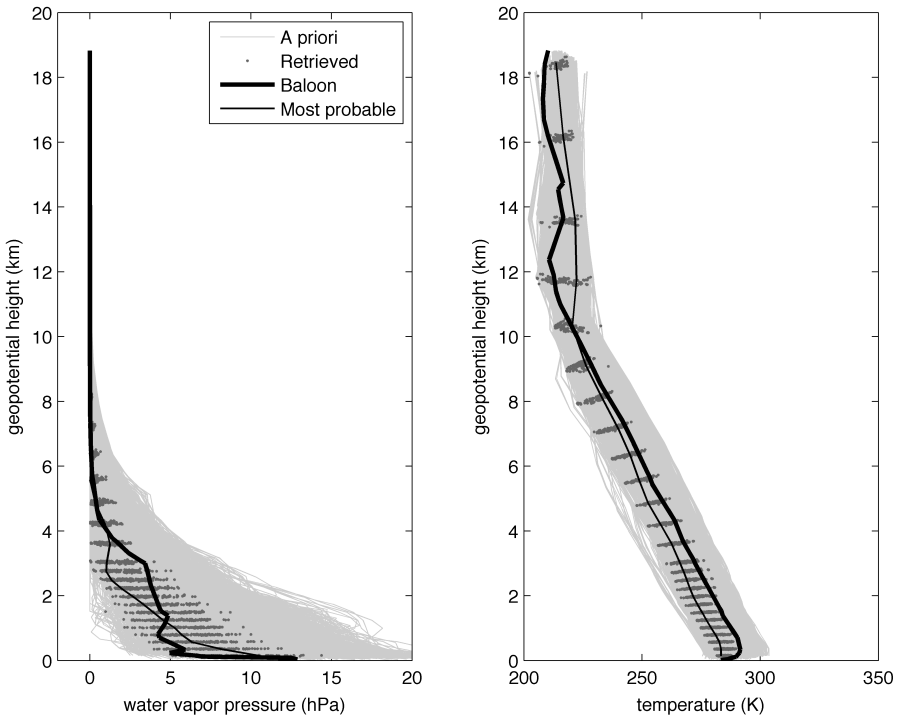


(b) Retrieval result using all measurements ($\tilde{P}_0$, $\tilde{T}_0$, $e_{\tilde{w}0}$, $Z\tilde{P}D$)

Figure 4.19: Same as Fig. 4.16 but date is 2011/04/14; time: 00:00 UTC and station is medi (Medicina).



(a) Retrieval result using only surface measurements ($\tilde{P_0}$, $\tilde{T_0}$, $e_{\tilde{w0}}$)



(b) Retrieval result using all measurements ($\tilde{P_0}$, $\tilde{T_0}$, $e_{\tilde{w0}}$, $Z\tilde{P}D$)

Figure 4.20: Same as Fig. 4.16 but date is 2011/04/11; time: 00:00 UTC and station is medi (Medicina).

4.4 Retrieval of precipitable water

The retrieval of the PW value has been obtained for each of the $\mathbf{T}$ and $\mathbf{e_w}$ (and $\mathbf{P}$)[10] profiles, using the relation derived from Eq. 1.38:

$$PW = \frac{1}{\rho} \int_0^{TOT} \rho_w \, dz \tag{4.35}$$

where ρ is the density of the water the integral value can be approximated as the finite summation of all the values of ρ_w at all profile levels, calculated by a formula derived by Eq. 1.26, i.e:

$$\rho_{wi} = \frac{e_{wi}}{T_i} \frac{M_w}{RZ_w} \tag{4.36}$$

The choice of a reasonable distribution scheme allows the comparison between the PW values associated to the retrieved profiles and their prior distribution, as well as the measured value given from balloon observations. This confirms that the results classified as "good" by visual inspection actually produce a more informative posterior PW distribution, whose most probable value is generally closer to the balloon measurement. In the next figures the result of the PW computation applied to all retrievals for the first configuration (Figs. 4.21 – 4.25), and for the second one (Figs. 4.26 – 4.30) are reported, for all the selected case studies shown previously.

The shown probability distributions are a normalised histogram applied to all the equiprobable profiles constituting the prior dataset (prior probability distribution) and the distribution resulting from the weights assigned to each of the selected profiles (posterior distribution). In most cases the exploitation of information derived from ground measurements, including the GNSS delay, changes dramatically the distribution. Since the posterior distribution assigns high probability to PW values with low prior probability and vice versa, this situation can be considered highly informative (see § 5.1 for more details).

The performances are sensibly better when using the coefficients calculated with only local data, as expected. However even in case of coefficients valid for the whole area, the estimated values of precipitable water are close to measured ones, although with a widening in the distribution. More details will be provided when a more detailed analysis of the values of precipitable water calculated for the full year 2011 will be made (§ 5.2).

A poor result is shown for the Ajaccio station (Fig. 4.28) where after the retrieval process the trimodal posterior probability distribution selects profiles with a two low water vapour content. The value of precipitable water is consequently underestimated

[10] In this work the results of the retrieval of pressure values in function of height are not shown, simply because they are less significant. In fact the variability of pressure in function of height very little differs from the standard profile of pressure. However, it is important to remember that for each profile in function of pressure isobaric levels are given Temperature, WV pressure and geopotential height (the latter used to rescale temperature and WV pressure in function of height).

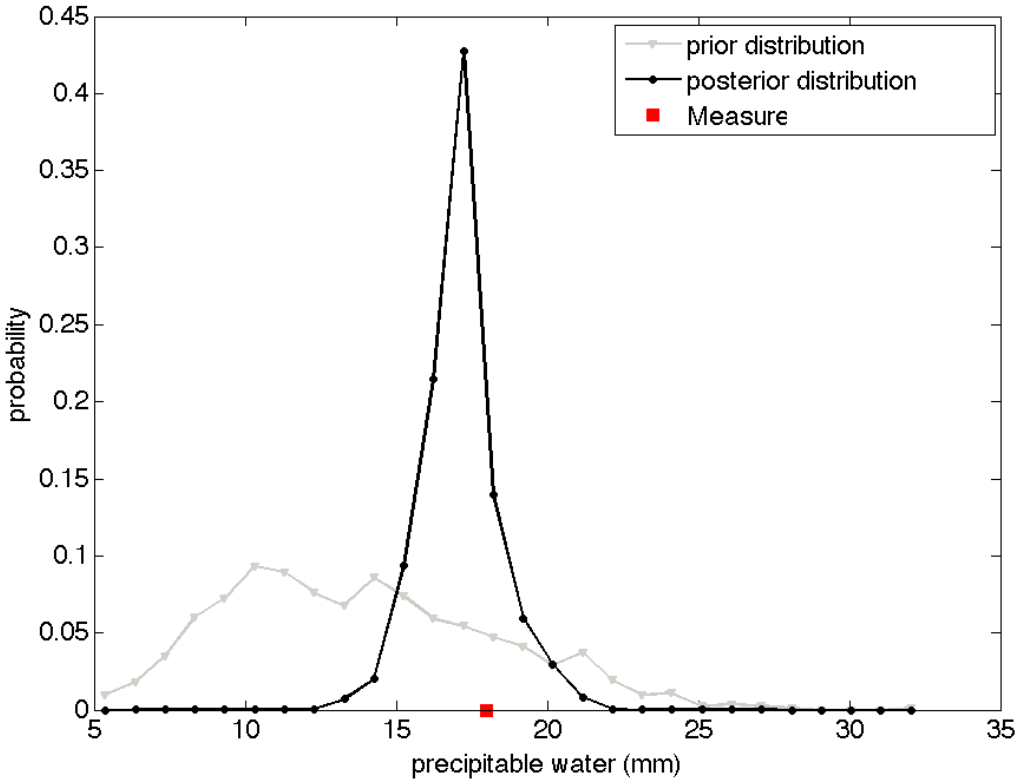


Figure 4.21: Example of posterior distribution for PW compared to the prior distribution and the value measured by balloon on the same day and time. Retrieval using locally valid coefficients (test one). Date: 2011/02/01 time: 00:00 UTC. Station is cagl (Cagliari).

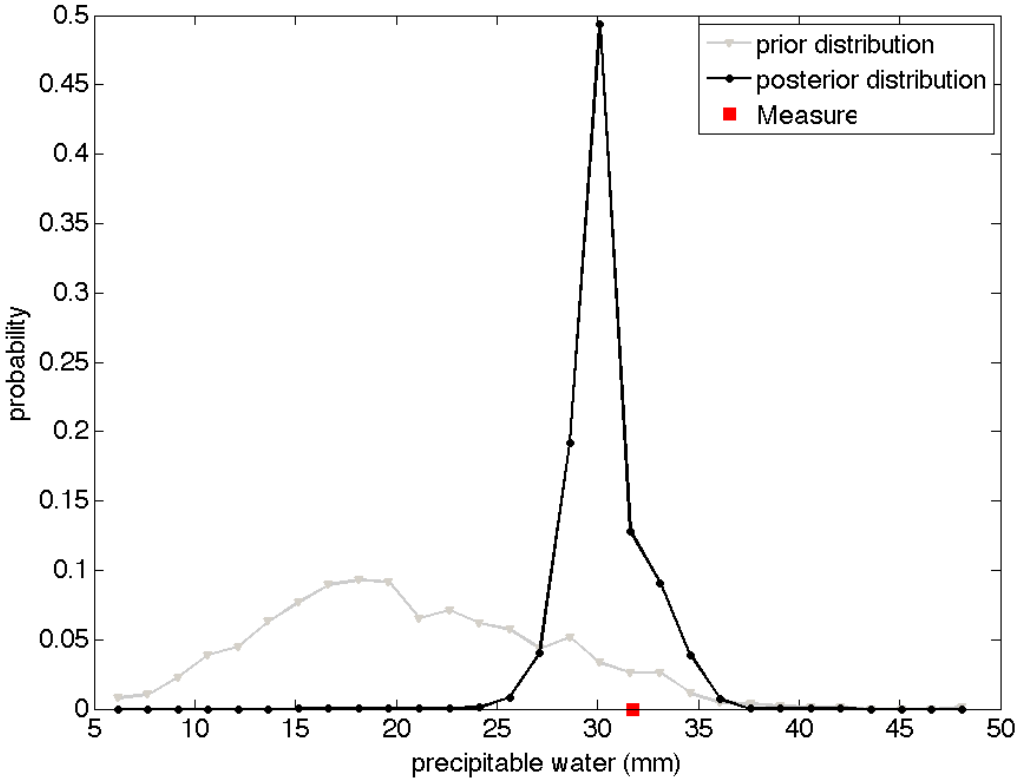


Figure 4.22: Same as Fig. 4.21 but for date: 2011/10/07 time: 12:00 UTC. Station is cagl (Cagliari).

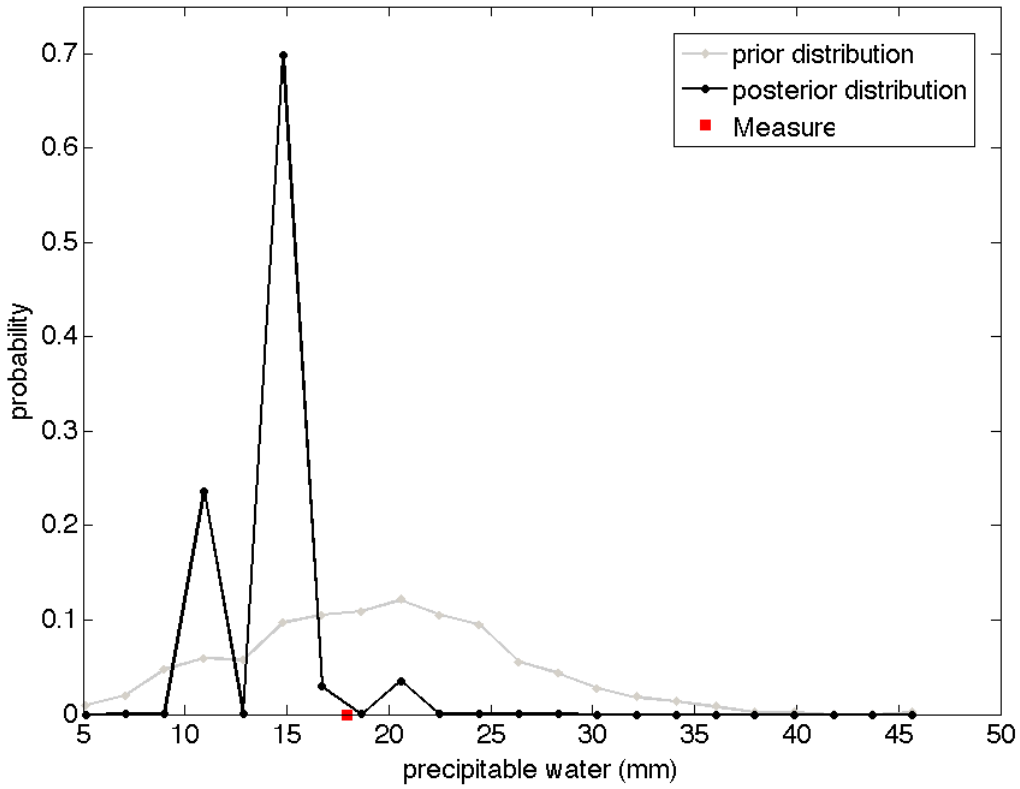


Figure 4.23: Same as Fig. 4.21 but for date: 2011/10/07 time: 12:00 UTC. Station is ajac (Ajacc).

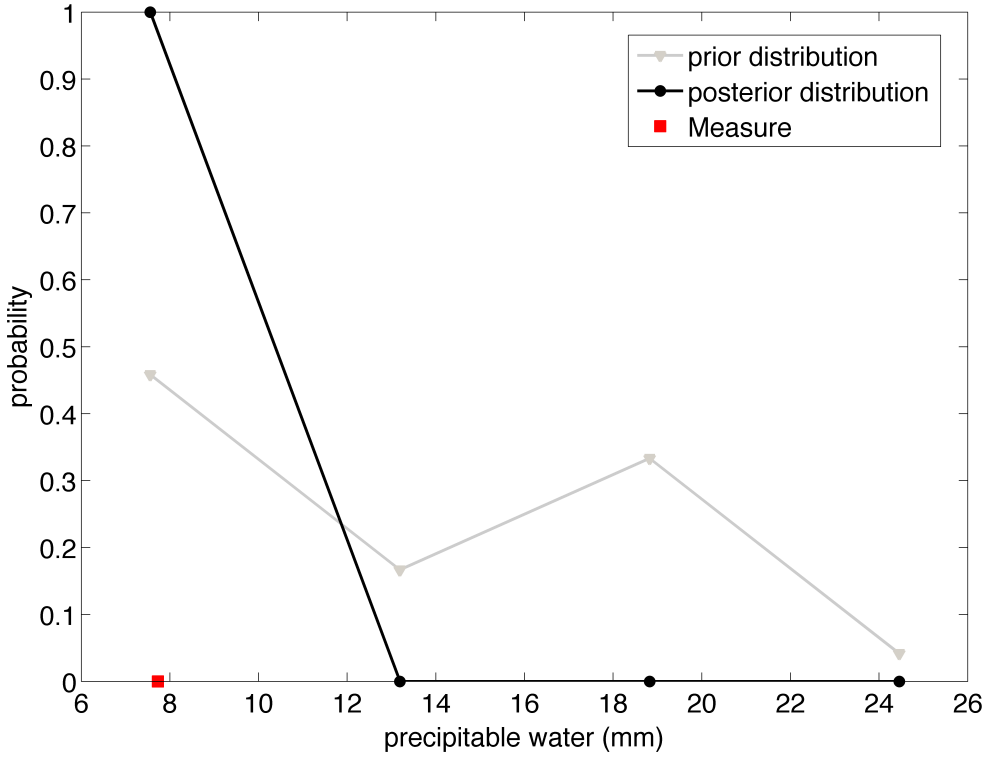


Figure 4.24: Same as Fig. 4.21 but for date: 2011/04/14 time: 00:00 UTC. Station is medi (Medicina). Note that in this case the posterior distribution is exactly one for only one value and zero elsewhere. In fact in this case only one profile is selected.

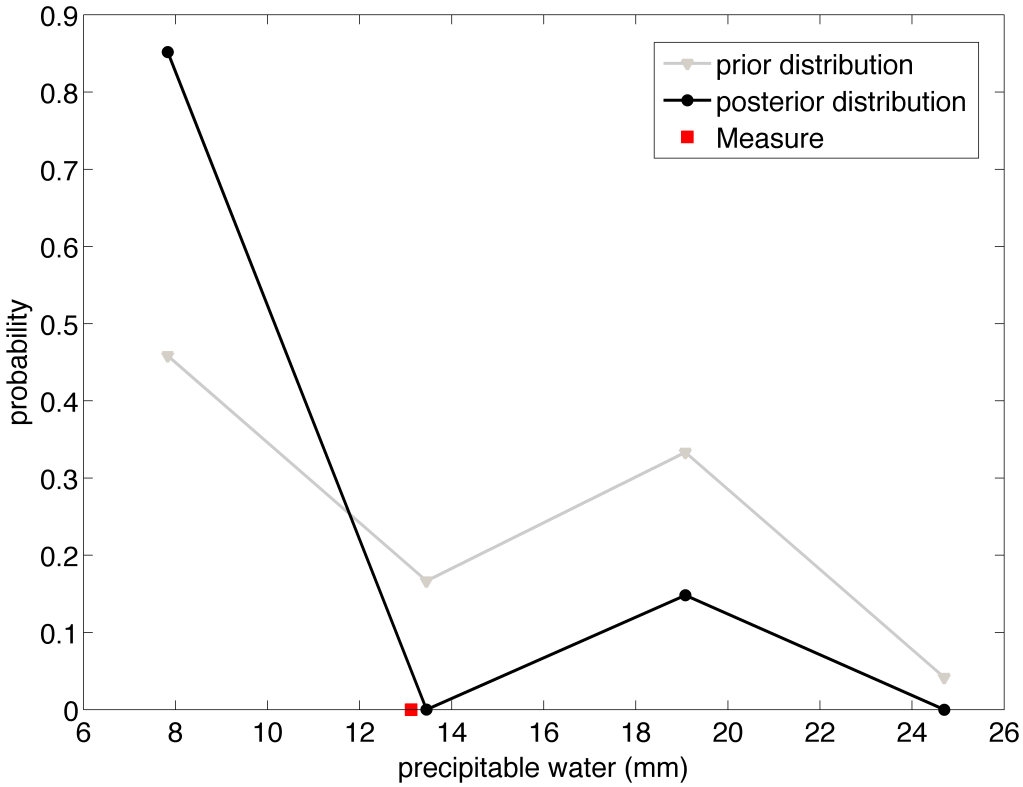


Figure 4.25: Same as Fig. 4.21 but for date: 2011/04/11 time: 00:00 UTC. Station is medi (Medicina).

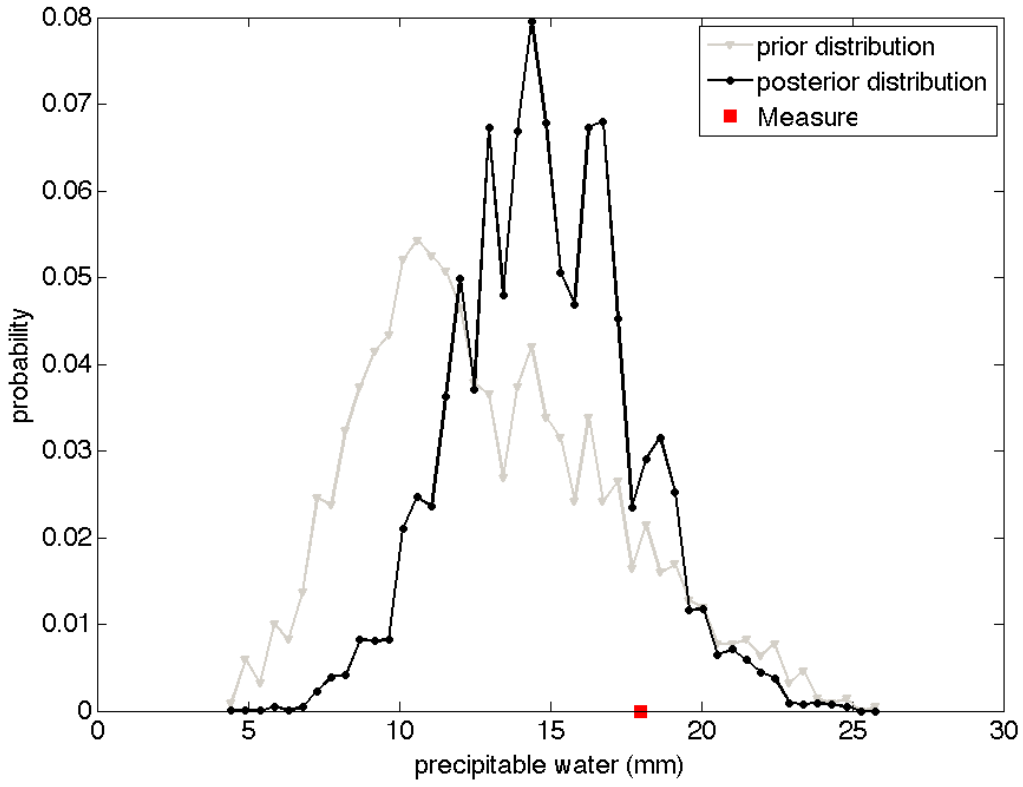


Figure 4.26: Example of posterior distribution for PW compared to the prior distribution and the value measured by balloon on the same day and time. Retrieval using whole area valid coefficients (test two). Date: 2011/02/01 time: 00:00 UTC. Station is cagl (Cagliari).

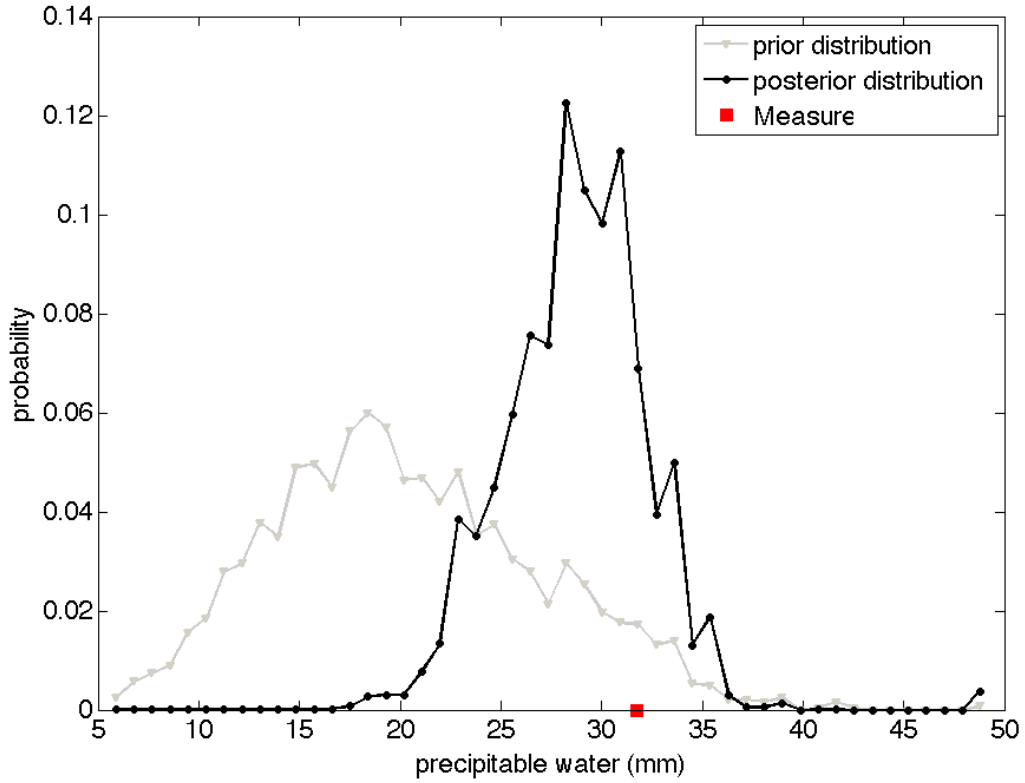


Figure 4.27: Same as Fig. 4.26 but for date: 2011/10/07 time: 12:00 UTC. Station is cagl (Cagliari).

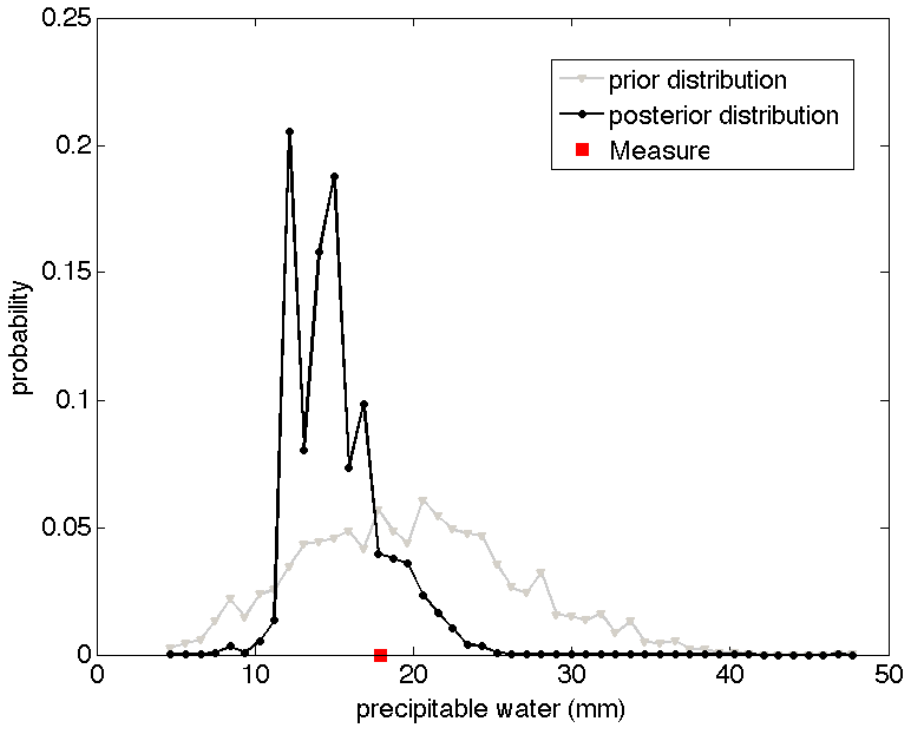


Figure 4.28: Same as Fig. 4.26 but for date: 2011/10/07 time: 12:00 UTC. Station is ajac (Ajacc).

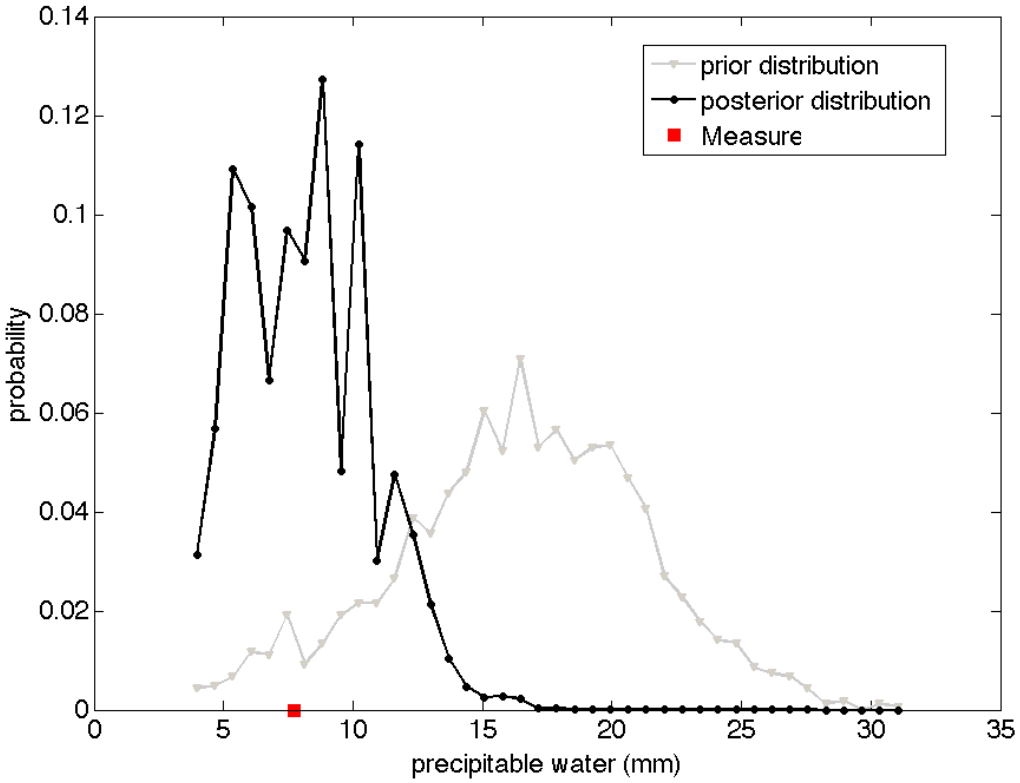


Figure 4.29: Same as Fig. 4.26 but for date: 2011/04/14 time: 00:00 UTC. Station is medi (Medicina).

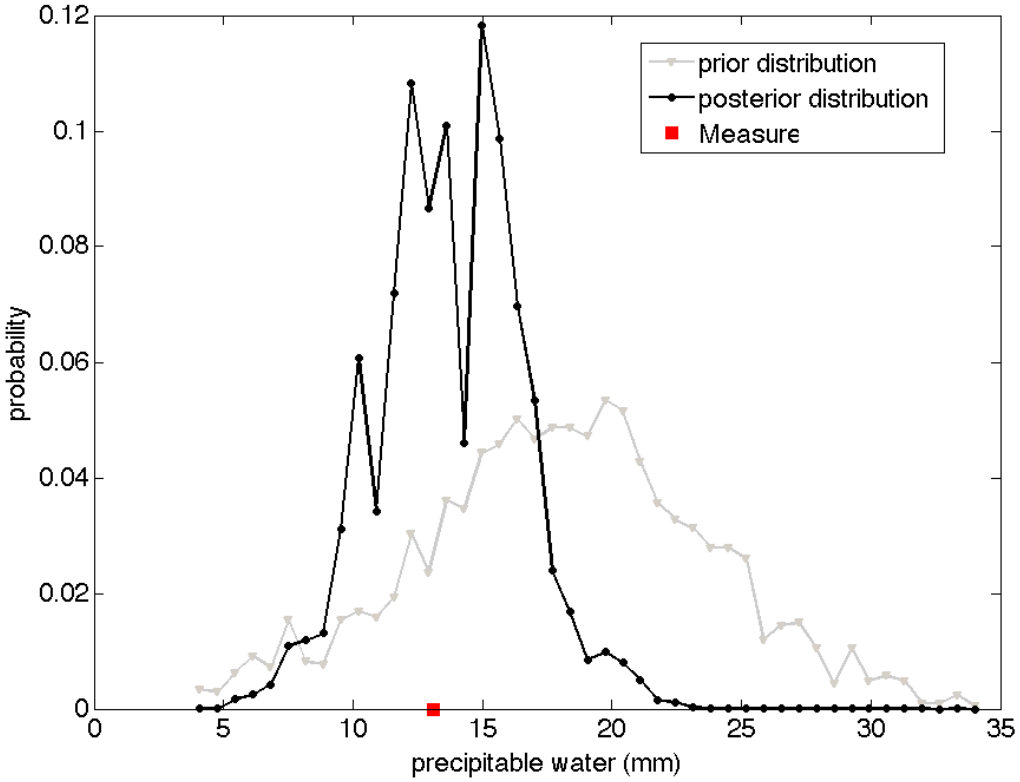


Figure 4.30: Same as Fig. 4.26 but for date: 2011/04/11 time: 00:00 UTC. Station is medi (Medicina).

for the causes previously discussed (see § 4.3.3).

A comment must be made to the images related to the Medicina (medi) test site. In the first test configuration the shape of the distribution is distorted by the low number of selected profiles, which becomes even only one (as in the case of Fig. 4.25) , resulting in a probability exactly equal to one. The results of retrieval drastically change, when switching from the coefficients of local validity to those valid on the whole area, because of the increased number of data in the prior dataset , which derives from collecting data from the wider area of the 4 neighbors pixels (see Fig. 4.29 and Fig. 4.30).

Chapter 5

Validation vs balloon data for three test sites

5.1 Analysis of entropy

The presented approach assumes the collected prior dataset as an enough representative system of mutually exclusive events. All the atmospheric states of the prior dataset are considered equiprobable (when no constraints are applied). When the measurement vector is available the posterior probability is used to select the atmospheric states. Each of the resulting finite schemes is characterised by an uncertainty degree, which is maximum when all the states have the same probability, and close to the minimum when just few similar states have a not negligible probability[1]. It is very useful to compute the changes of the uncertainty when passing from one scheme to an other one. The quantity of uncertainty can be assessed by analysing the entropy of the finite scheme, which allows the quantitative assessment of the information gain resulting from the addition of the observations, and in particular of the GNSS troposheric delay, as explained below.

For a generic finite scheme of events $A_1, A_2, ..., A_M$ with assigned probabilities $Pr_1, Pr_2, ..., Pr_M$ the entropy H is defined as (see for instance [63]):

$$H = \sum_{i=1}^{M} Pr_i \log \frac{1}{Pr_i} = -\sum_{i=1}^{M} Pr_i \log Pr_i \tag{5.1}$$

The entropy has its minimum value of $H = 0$ if only one event has the maximum probability value ($Pr = 1$) and consequently all others have probability $Pr = 0$. This is the case of minimum uncertainty of the system, in any other case the entropy has a positive value.

In the proposed approach if the prior probability for the M equally likely states $\mathbf{x}_i$

[1]The minimum value of uncertainty occurs when only one state has a not null probability which obviously results equal to one, and that leads to a deterministic system.

is $Pr(\mathbf{x}_i) = \frac{1}{M}$ the highest entropy value is defined as:

$$H_{prior} = -\sum_{i=1}^{M} Pr(\mathbf{x}_i) \log Pr(\mathbf{x}_i) = \log M \tag{5.2}$$

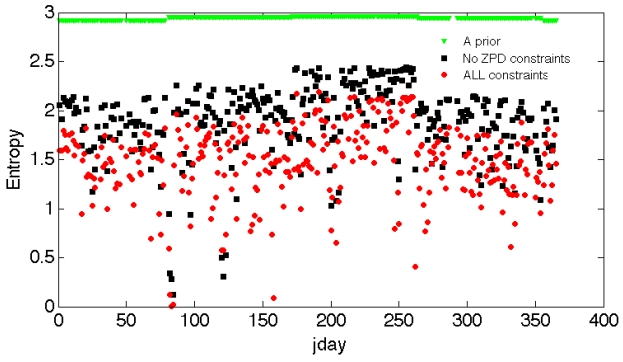
The addition of pieces of information given by the observations $\tilde{\mathbf{y}}$ changes it to:

$$H(\tilde{\mathbf{y}}) = -\sum_{i=1}^{M} Pr(\mathbf{x_i}|\tilde{\mathbf{y}}) \log Pr(\mathbf{x_i}|\tilde{\mathbf{y}}) \leq \log M \tag{5.3}$$

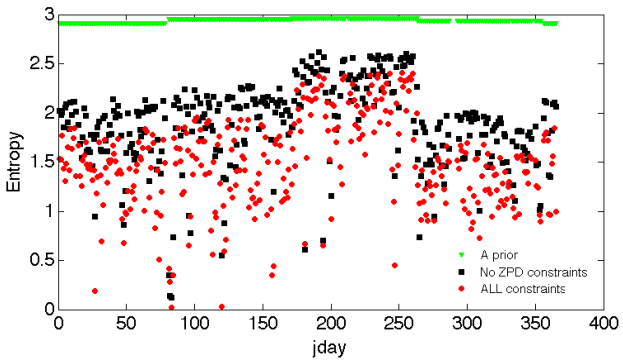
The difference $logM - H(\tilde{\mathbf{y}})$ quantifies the information gain due to the observations, which are as more informative as this difference increases. Analogously the difference $H(\tilde{\mathbf{y}}_\mathbf{0}) - H(\tilde{\mathbf{y}})$ indicates the information gain coming from the addition of the GNSS troposheric delay measurement to the basic set of ground meteorological observations $\tilde{\mathbf{y}}_\mathbf{0} = [\tilde{P}_0,\ \tilde{T}_0,\ \tilde{e}_{w0}]$.

For the selected sites the results of entropy computation are represented in the following figures, when the same probability is assigned to all the dataset and when in the retrieval process the ground measurements and the full set of measurement are considered. The first configuration of coefficients developed to a specific station are relative to Figs. 5.1 – 5.3, and the second test results of coefficients available for the full area are represented in Figs. 5.4 – 5.6.

The results of entropy computation confirm the decrease of the entropy values

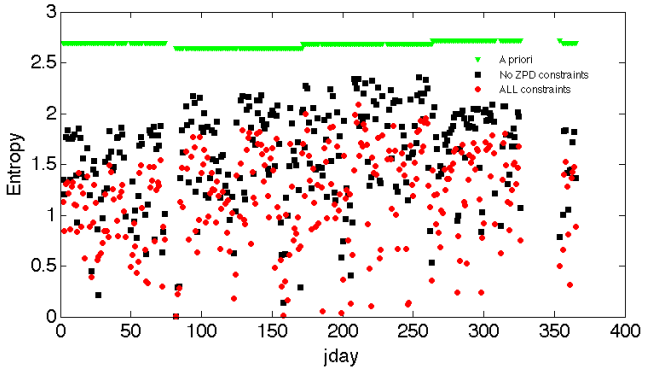


(a) 00:00 UTC

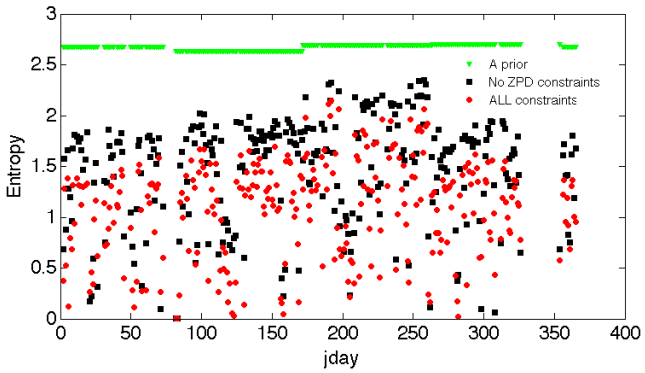


(b) 12:00 UTC

Figure 5.1: Entropy averaged values of retrieved profiles varying the constraint measurements. First test configuration is shown. Period full 2011 year. Station is cagl (Cagliari).

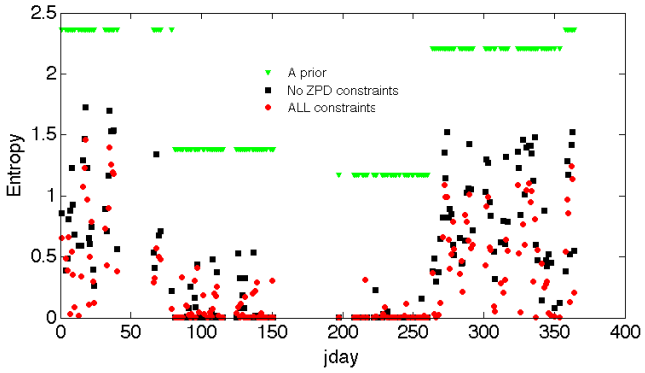


(a) 00:00 UTC

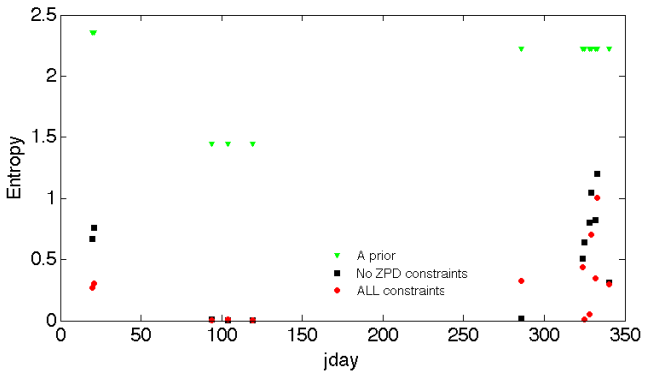


(b) 12:00 UTC

Figure 5.2: Same as Fig. 5.1, but for station of ajac (Ajaccio).

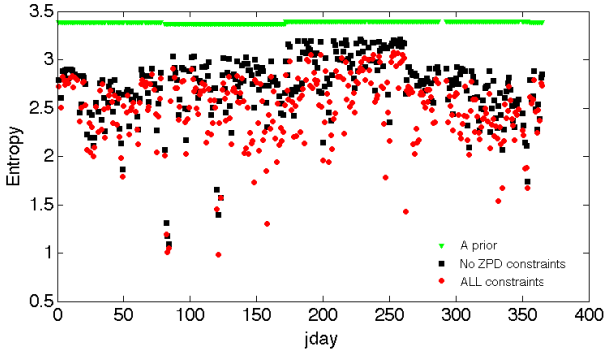


(a) 00:00 UTC

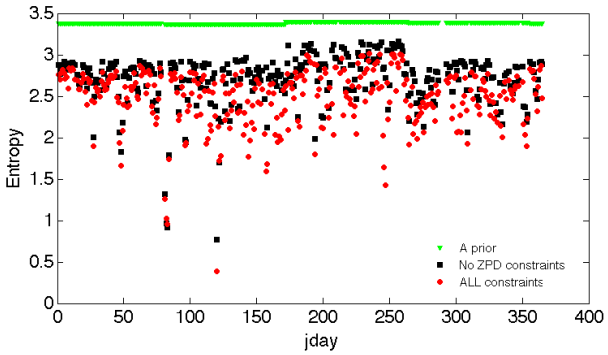


(b) 12:00 UTC

Figure 5.3: Same as Fig. 5.1, but for station of medi (Medicina).

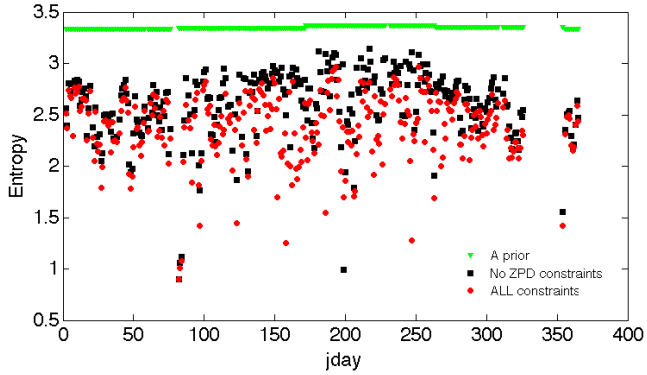


(a) 00:00 UTC

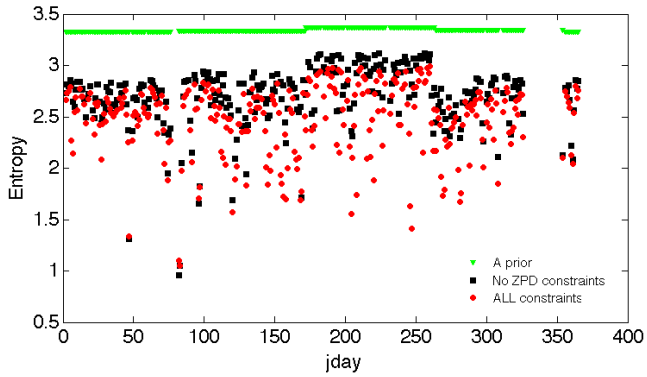


(b) 12:00 UTC

Figure 5.4: Entropy averaged values of retrieved profiles varying the constraint measurements. Second test configuration is shown. Period full 2011 year. Station is cagl (Cagliari).

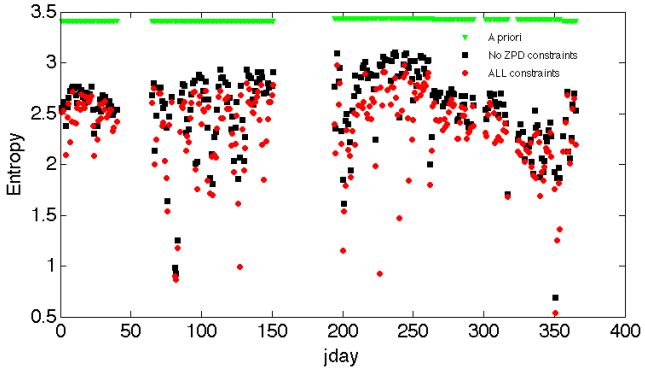


(a) 00:00 UTC

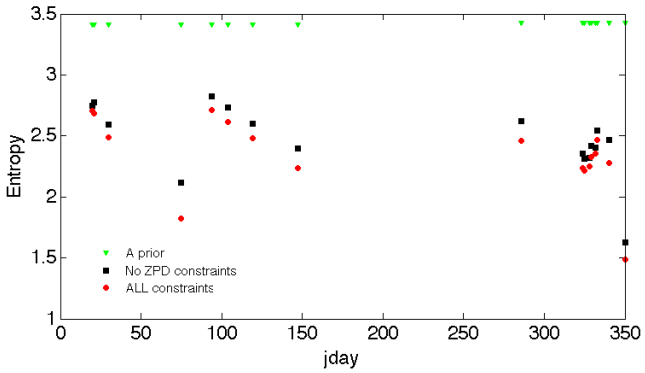


(b) 12:00 UTC

Figure 5.5: Same as Fig. 5.4, but for station of ajac (Ajaccio).



(a) 00:00 UTC



(b) 12:00 UTC

Figure 5.6: Same as Fig. 5.4, but for station of medi (Medicina).

when the number of measurements increases. Moreover the selectivity of the second method is lower as the normal distributions are wider, resulting in a lower information gain introduced by the use of measurements.

It must be noted that the number of profiles constituting the prior dataset drastically increases when we pass from data accumulated only for one MERRA pixel for a retrieval applicable only on a single station, to four MERRA pixels, to get the validity over the whole target area, and consequently the degree of indeterminacy of the system state (i.e. the entropy) is higher. The number of profiles constituting the prior dataset for each station, each season, each available time of the day (00:00 UTC and 12:00 UTC) and for both test configurations, is shown in Table 5.1.

Knowing the number of profiles the analysis of the entropy decrease for the shown cases is clearer. It is obvious that a really poor (i.e. incomplete) prior dataset, although related to a strong entropy reduction, is often the cause of a wrong reconstruction of the atmospheric profile, as shown in more details below, with the selected atmospheric state very different from the real one. As the increment of the number of the prior states have been obtained by enlarging the area of interest (i.e. considering a number of contiguous pixels), the effect of an excessive number of number of profiles in the dataset can be the opposite, i.e. an excessive number of profiles is related to the risk of including a series of atmospheric states that can not apply in the site of analysis. A

Station	Method	Profiles number by season			
		Spring	Summer	Autumn	Winter
Cagliari @ 00	**1 MERRA PIXEL**	894	920	878	828
	4 MERRA PIXELS	2351	2493	2481	2443
Cagliari @ 12	**1 MERRA PIXEL**	893	919	873	820
	4 MERRA PIXELS	2328	2479	2434	2403
Ajaccio @ 00	**1 MERRA PIXEL**	431	496	504	475
	4 MERRA PIXELS	2202	2324	2256	2172
Ajaccio @ 12	**1 MERRA PIXEL**	431	496	504	475
	4 MERRA PIXELS	2186	2332	2232	2141
Medicina @ 00	**1 MERRA PIXEL**	24	15	163	231
	4 MERRA PIXELS	2570	2715	2689	2579
Medicina @ 12	**1 MERRA PIXEL**	28	13	167	226
	4 MERRA PIXELS	2571	2712	2677	2577

Table 5.1: Number of profiles constituting the prior dataset.

reasonable compromise has been found in this work considering a neighborhood of up to 4 pixels of the MERRA model, resulting on an area of about 110 km $\times$ 110 Km for the selected area.

The entropy values computed on the basis of the number of retrieved profiles and their probability could not give any information about the single isobaric level values. Moreover a large range of values relative to one vertical level can not be considered in the same manner as a limited range of values of another level. For example the water vapour content at lower vertical levels can cover a large dynamic of values, while in the upper levels is close to zero in any case. For this reason we have performed the entropy analysis level by level, limiting to the first 27 levels starting from the ground (i.e. all the troposphere including also the tropopause), which are the most significant especially for the content of water vapour.

For each considered variable V of the vector state $\mathbf{x}$ (one of h, T, e_w) at each j-th vertical pressure level the range of values, given by the prior dataset values (at this level), is limited by minimum and maximum values (V_{min} and V_{max}). This interval is divided in N bins.

As previously discussed for the prior dataset the probability of each value V_{ij} of the considered variable for the i-th profile at the j-th level is the same and is equal to $Pr(V_{ij}) = \frac{1}{M}$. If for each bin there is the occurrence of n_k values the cumulated pressure at this bin is $\frac{n_k}{M}$ and the computation of the entropy for the j-th pressure level is:

$$H_{V_j} = -\sum_{k=1}^{N} \frac{n_k}{M} log \frac{n_k}{M} \tag{5.4}$$

When the retrieval is applied a weight is assigned to each profile ($Pr(\mathbf{x}_i)$) and the same weight is valid also for the values of all the vertical levels V_{ij}: $Pr(V_{ij}) = Pr(\mathbf{x}_i)$.

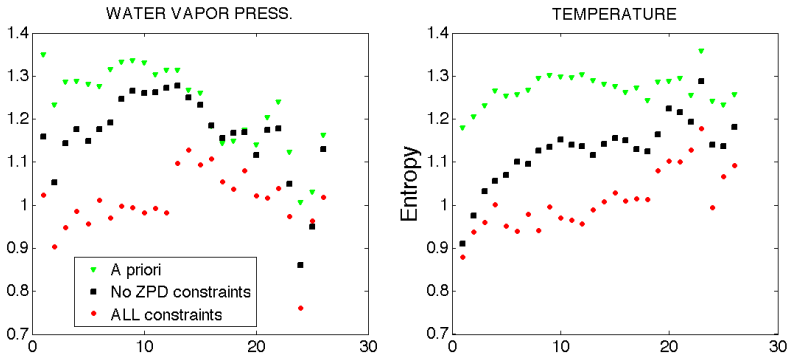
The entropy is thus computed as:

$$H_{V_j}(\tilde{\mathbf{y}}) = -\sum_{k=1}^{N}\left(\sum_{i=1}^{n_k} Pr(V_{ij}|\tilde{\mathbf{y}})\right) log\left(\sum_{i=1}^{n_k} Pr(V_{ij}|\tilde{\mathbf{y}})\right) \tag{5.5}$$

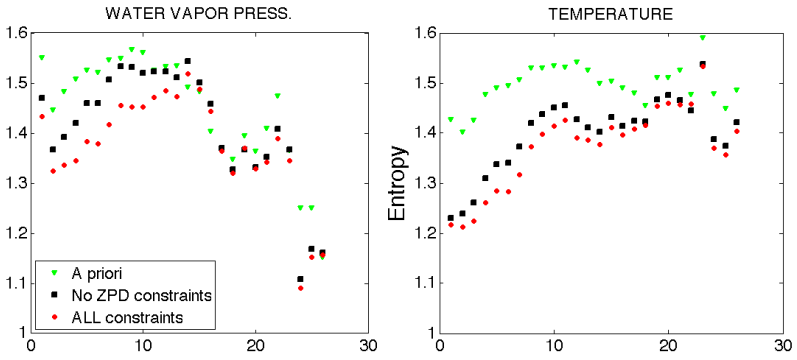
where $\sum_{i=1}^{n_k} Pr(V_{ij}|\tilde{\mathbf{y}})$ is the sum of the weights of the n_k profiles whose values V_{ij} fall within the k-th bin. If the observable vector does not include the measurement of ZPD the entropy $H_{V_j}(\tilde{\mathbf{y}}_\mathbf{0})$ is usually larger, a confirmation of the increase of the information content when introducing the ZPD measurement constraint.

Some entropy computation results relative to the same previously analysed case studies are depicted in the following figures (Figs. 5.7 – 5.11). Both algorithm configuration are shown.

Note that in the shown cases the entropy of the prior distribution in some vertical levels is lower than those of the posterior distribution. This is not surprising as it can come out due to a redistribution of the weights when passing from Eq. 5.4 to Eq. 5.5. This occurs only in some individual cases, as those shown in the following figures, but, as shown below, the average values prove a decrease of the entropy values when the number of measurements (or the information content) increases.

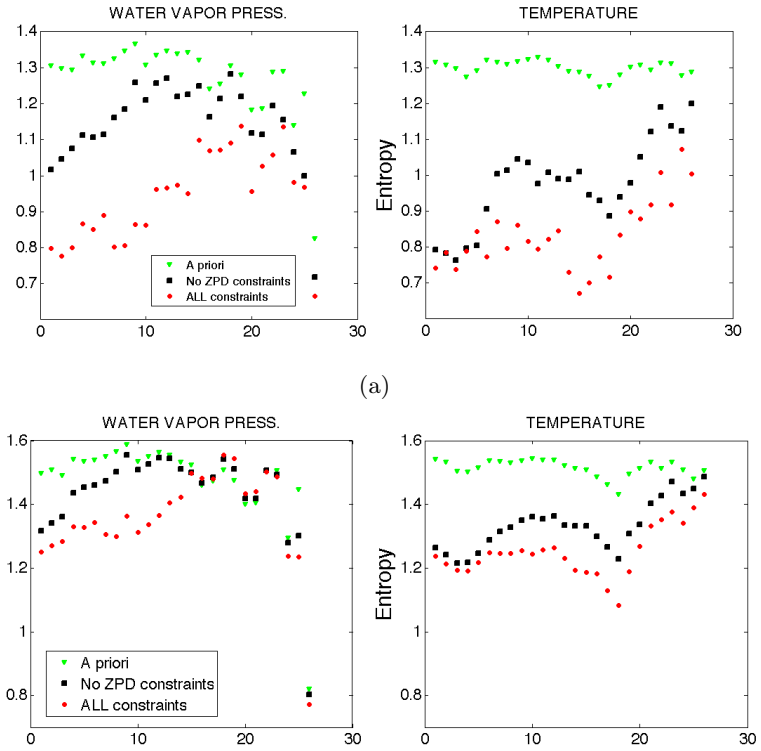


(a)



(b)

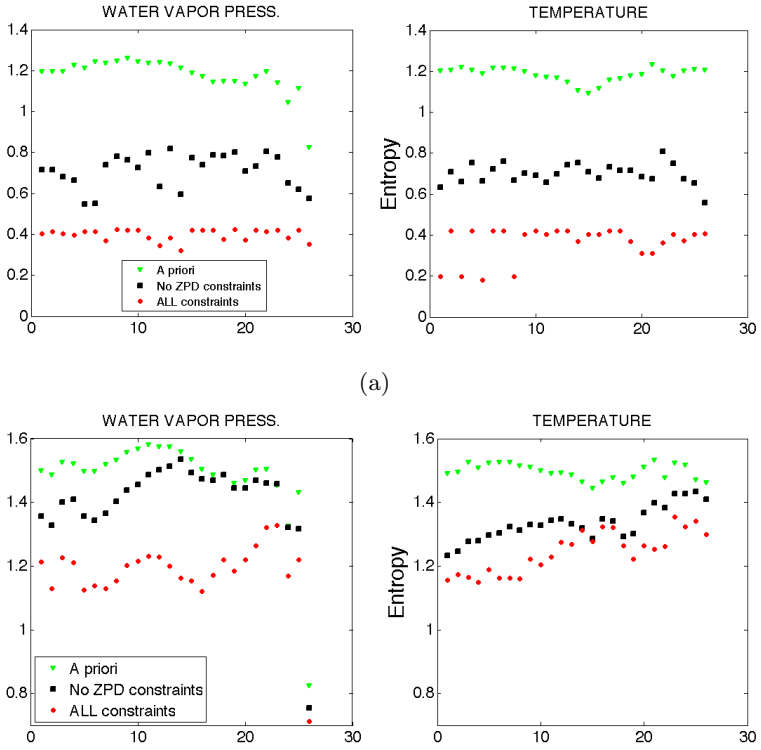
Figure 5.7: Entropy of WV pressure and temperature for the first 27 atmospheric levels considering the prior distribution, using only the basic ground measurement and including also the GNSS ZPD measurement. Both method are shown: the first configuration (a) and the second test configuration (b). Date: 2011/02/01 time: 00:00 UTC. Station is cagl (Cagliari).



(a)

(b)

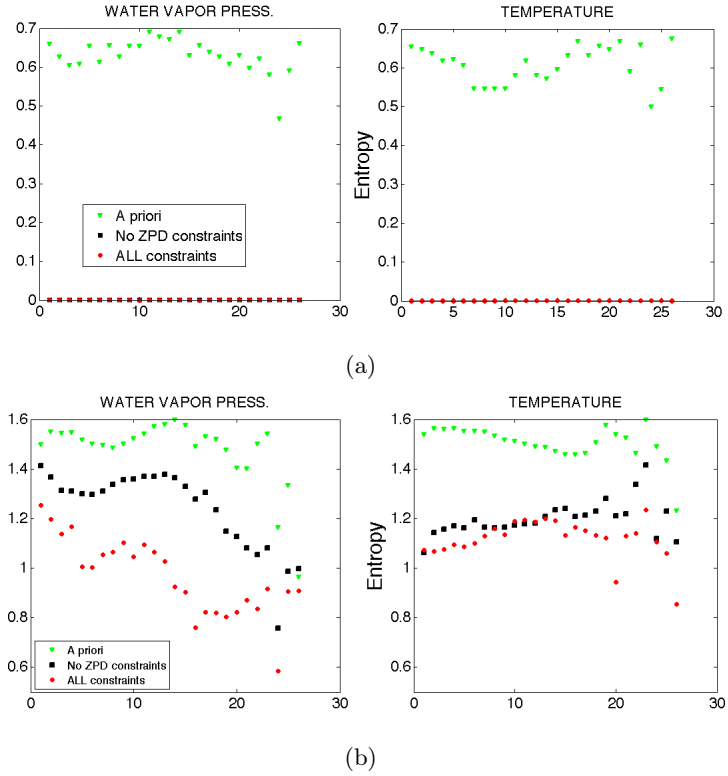
Figure 5.8: Same as Fig. 5.7, but for date: 2011/10/07; time: 12:00 UTC and station of cagl (Cagliari).



(a)

(b)

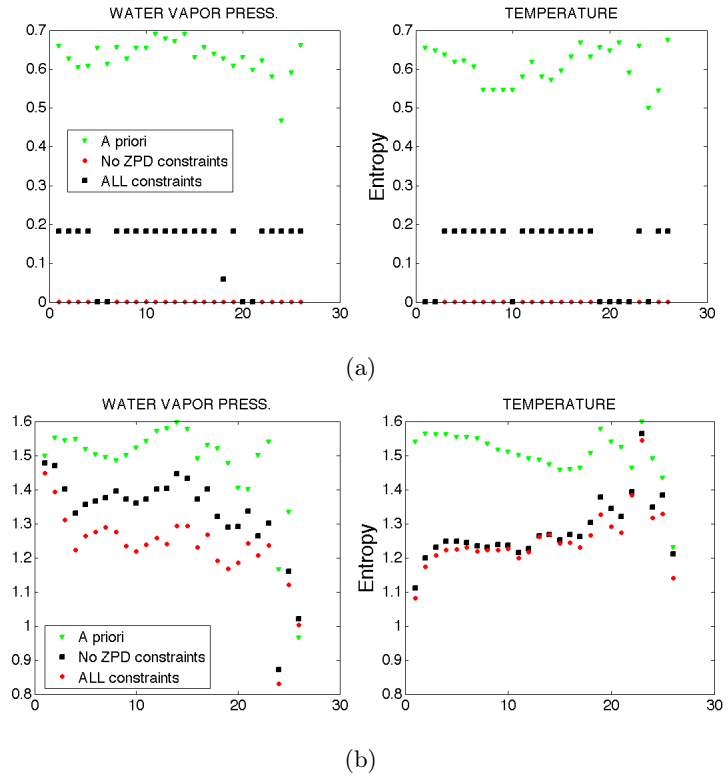
Figure 5.9: Same as Fig. 5.7, but for date: 2011/10/07; time: 12:00 UTC and station of ajac (Ajaccio).



(a)

(b)

Figure 5.10: Same as Fig. 5.7, but for date: 2011/04/14; time: 00:00 UTC and station of medi (Medicina). Note that as shown in § 4.3.3, only one profile is selected from the retrieval (a) with probability 1, and entropy 0.



(a)

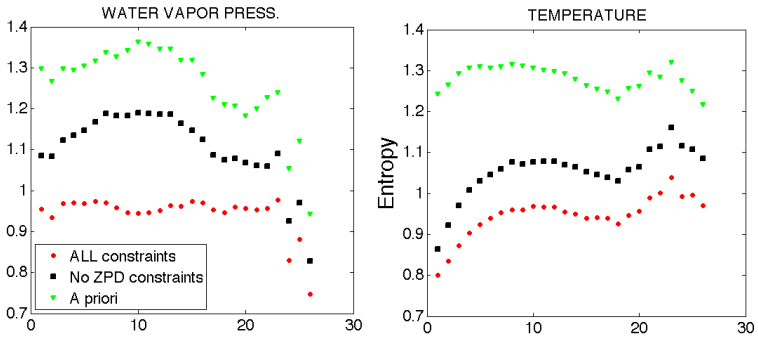
(b)

Figure 5.11: Same as Fig. 5.7, but for date: 2011/04/11; time: 00:00 UTC and station of medi (Medicina). Note that as shown in § 4.3.3, only two profiles are selected from the retrieval (a) with high probability, and consequently low entropy.

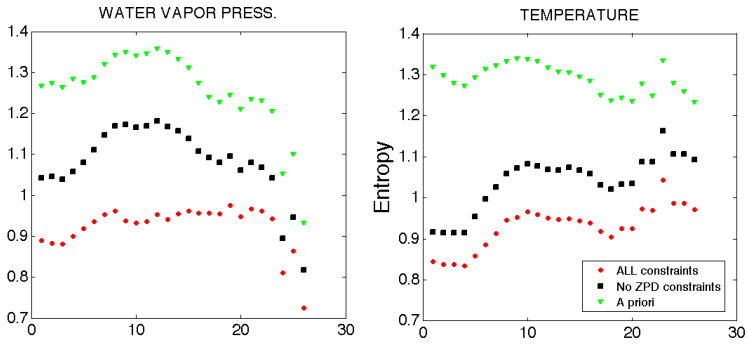
The first configuration uses the spatial information directly in the prior distribution determination, and consequently allows to have a less entropic system. Moreover the application of the algorithm with the coefficient set customised for the application site, results in a decrement of the entropy values also when only the ground observations are used. The ground observations allow to give a lot of information about the values of temperature (the informative content is remarkable near the surface), but also about the WV content of the lower vertical levels. The ZPD measurement adds a little information about temperature in the intermediate vertical levels (in the first levels the entropy values for the temperature with or without the ZPD are very similar) and it is highly informative for the WV content at all vertical levels.

It can be noted that when the geographical area is expanded to 4 MERRA pixels, introducing a large number of profiles, the entropy of the prior distribution increases, as expected, but also the selective capability of the algorithm for the already mentioned reasons is lower. However an advantage, in the second configuration of the method, lies in the fact that, if it is not possible to apply the retrieval to a single MERRA pixel (first configuration), because of the lack of measures or of MERRA data, there are more chances to get satisfactory results, as it is extended to the four pixels surrounding the point of interest. One example is the case of Medicina, in which, for the first configuration, there are not enough data constituting the profiles dataset to be selected. In fact in the selected cases the retrieval process selects only one or two profiles both using only ground measurements constraints and including GNSS measurements. In addition the profiles selected by the retrieval are in some cases very different from the measured ones. This also leads to the singularity that the introduction of an additional constraint does not change the value of entropy which is very close or equal to zero (as in Fig. 5.10(a) and Fig. 5.11) . When considering the second configuration the scenario considerably changes and the results are comparable with those of the other stations.

As a more statistically significant result, which confirms the analysis carried out so far about entropy, below the annual (2011) averages of the entropy values calculated level by level starting from single dates values, for both times of the day (00:00 UTC and 12:00 UTC) and for both the configuration experiments (see Figs. 5.12 – 5.17). The considerations discussed above are also valid for the analysis of these figures, which confirm the monotonically descending entropy values as a function of the number of introduced constraints. Moreover, because the information content about temperature is mainly introduced by the ground set of measurements, the introduction of ZPD very slightly reduces the values of temperature entropy, but significantly reduces the entropy values of the water vapour content, in particular at vertical levels where the water vapour content is generally large. It is not surprising to note that the performance does not change significantly depending on the time of the day.

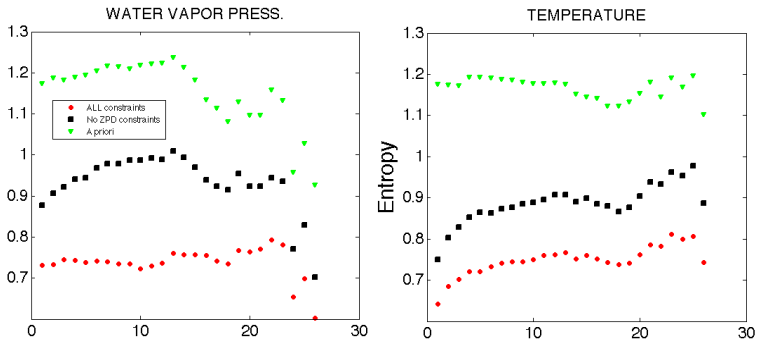


(a) 00:00 UTC

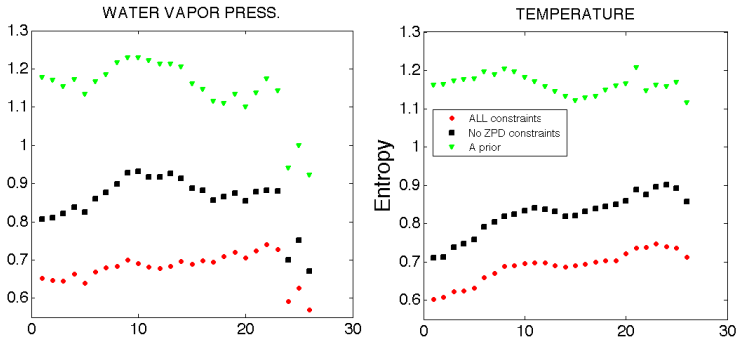


(b) 12:00 UTC

Figure 5.12: Annual (2011) average entropy of WV pressure and temperature for the first 27 atmospheric levels considering the prior distribution, using only the basic ground measurement and including also the GNSS ZPD measurement. First configuration. Station is cagl (Cagliari).

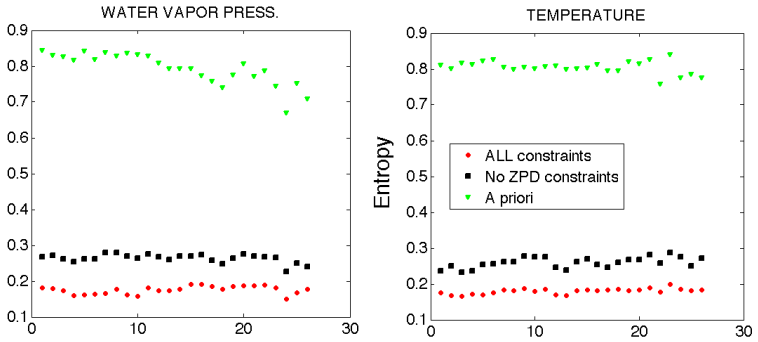


(a) 00:00 UTC

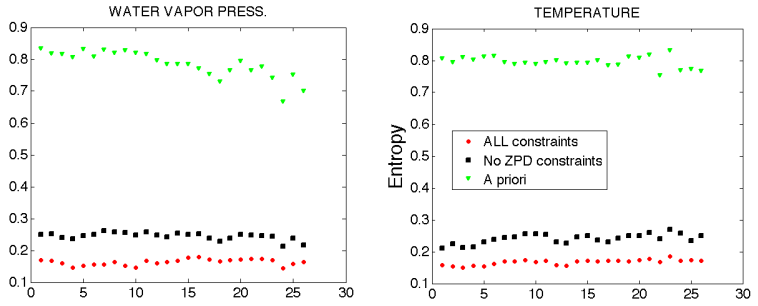


(b) 12:00 UTC

Figure 5.13: Same as Fig. 5.12 but for station of ajac (Ajaccio).

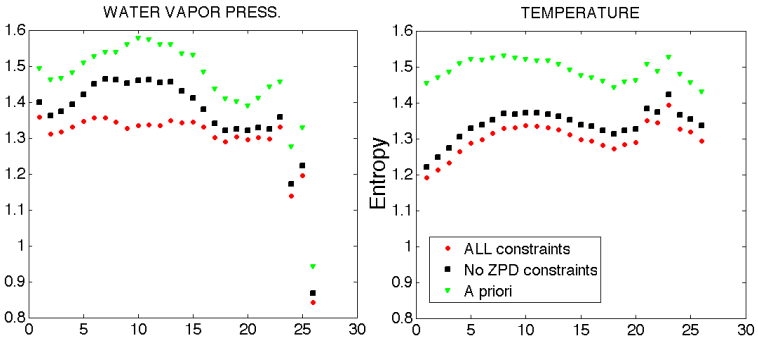


(a) 00:00 UTC

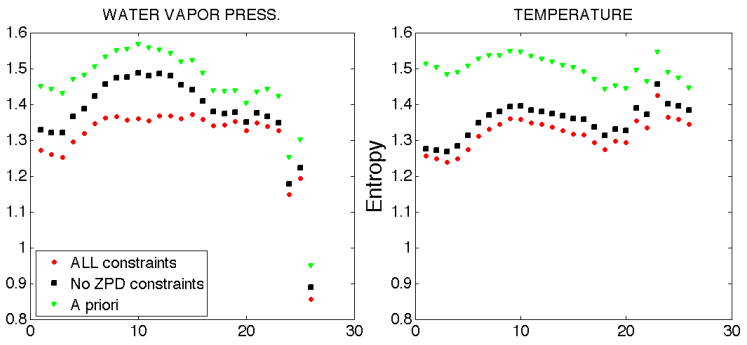


(b) 12:00 UTC

Figure 5.14: Same as Fig. 5.12 but for station of medi (Medicina).

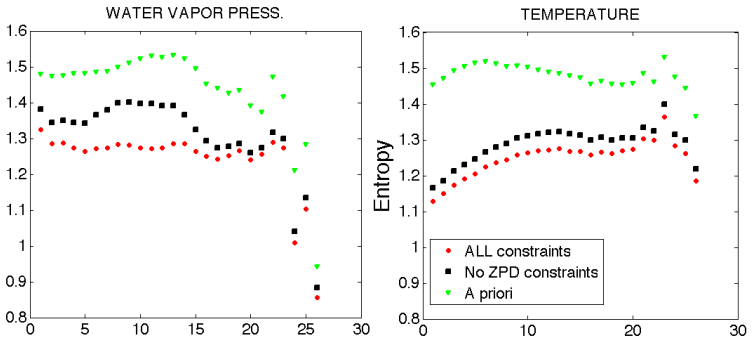


(a) 00:00 UTC

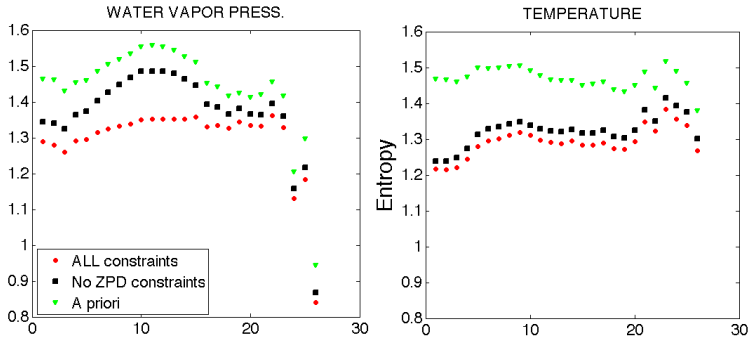


(b) 12:00 UTC

Figure 5.15: Annual (2011) average entropy of WV pressure and temperature for the first 27 atmospheric levels considering the prior distribution, using only the basic ground measurement and including also the GNSS ZPD measurement. Second test configuration. Station is cagl (Cagliari).

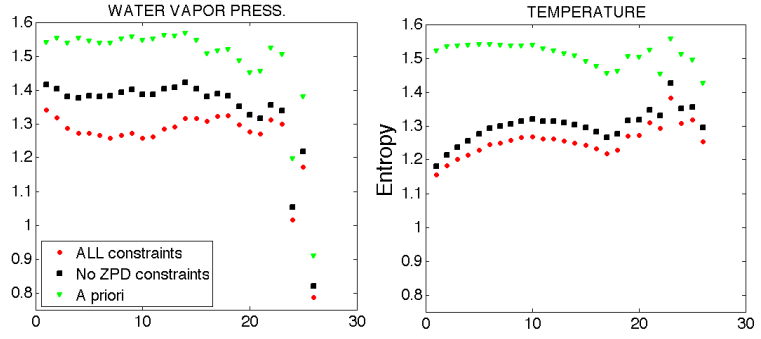


(a) 00:00 UTC

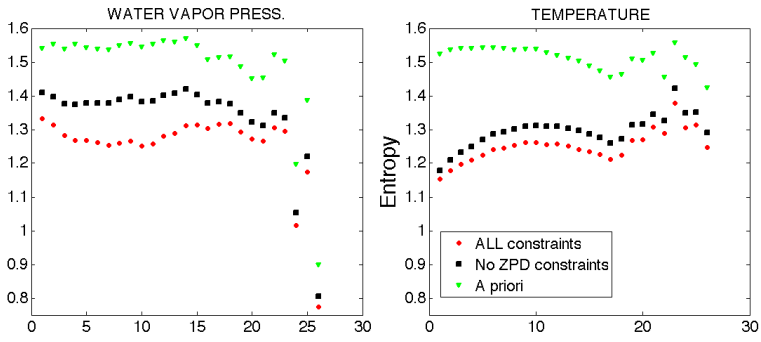


(b) 12:00 UTC

Figure 5.16: Same as Fig. 5.15 but for station of ajac (Ajaccio).



(a) 00:00 UTC



(b) 12:00 UTC

Figure 5.17: Same as Fig. 5.15 but for station of medi (Medicina).

In addition, for the temperature, the entropy values have an increment in function of height, confirming the higher correlation between the ground observations and the lower vertical atmospheric states variables. When at higher isobaric levels this correlation reduces, the spread of values, as well as entropy increases. In the case of temperature, when using only surface measurements but also when introducing the ZPD observation, entropy directly follows the trend of the prior entropy, level by level, although with lower values. For the case of water vapour, the introduction of the measured value of ZPD reduces the spread of entropy values within the isobaric levels, and the resulting trend can significantly differ from the prior entropy values.

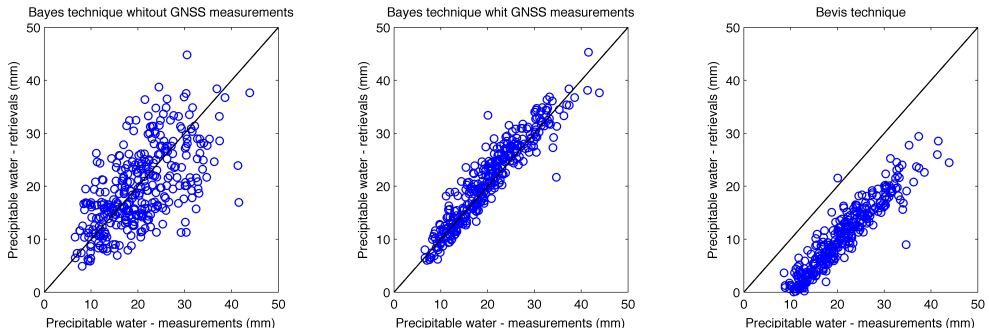
The analysis of entropy can not be used to evaluate the success of the retrieval, that instead derives from additional information about the uncertainty of the estimate. In fact, the information content is connected to the indeterminateness of a process and then to its uncertainty. To evaluate the accuracy it is necessary to introduce a further analysis, such as the estimation of parameters that provide information about the errors, for example the RMSE and BIAS, which will be discussed later.

5.2 PW retrieval

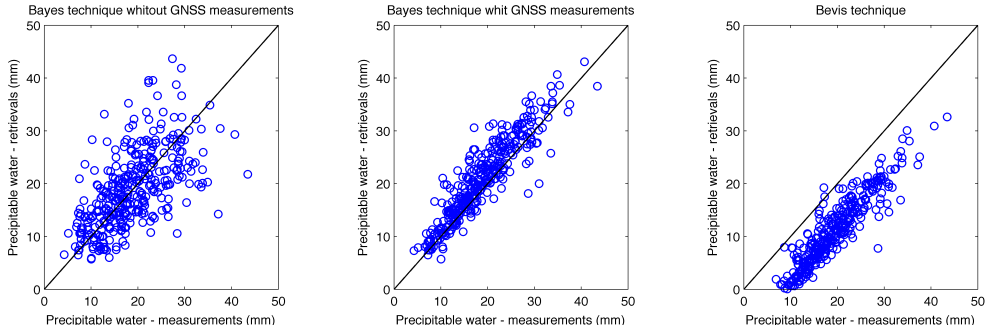
A simulation of operational precipitable water (PW) retrieval during a full year (2011), has been performed for the three tests sites using both configuration tests. The method used for the computation of PW has been described in § 4.4. As previously shown the knowledge of the ground measurements result has been assumed, mimicking two different situations: when only the basic ground observations (pressure, temperature and humidity) are available and when also the GNSS tropospheric delay is measured, to give the extended observable $\tilde{\mathbf{y}}$. Taking advantage of these pieces of information, the posterior probability is calculated for each state $\mathbf{x}_i$, and then associated to the corresponding IWV($\mathbf{x}_i$) or PW($\mathbf{x}_i$) value (see § 1.3 and § 1.3.3). The posterior distribution of PW is then easily determined, as well as its entropy and most probable value.

The introduction of the basic ground observations always produces a PW posterior distribution whose entropy is lower than the prior distribution entropy, as expected. More interesting, the additional use of the GNSS delay measurement further decreases the entropy in the large majority of the examined days, to suggest an intake of new useful information. However the decrease in entropy indicates only that the PW posterior distribution becomes less dispersed once the GNSS delay is introduced. In other words the GNSS PW estimate turns out more *precise*, but not necessarily be more *accurate*. To verify the accuracy a direct comparison with a standard of reference, like the PW values calculated by the balloon data are needed.

In next figures (Figs. 5.18 – 5.23) the scatter plots of PW most probable values versus the balloon measured values are shown.

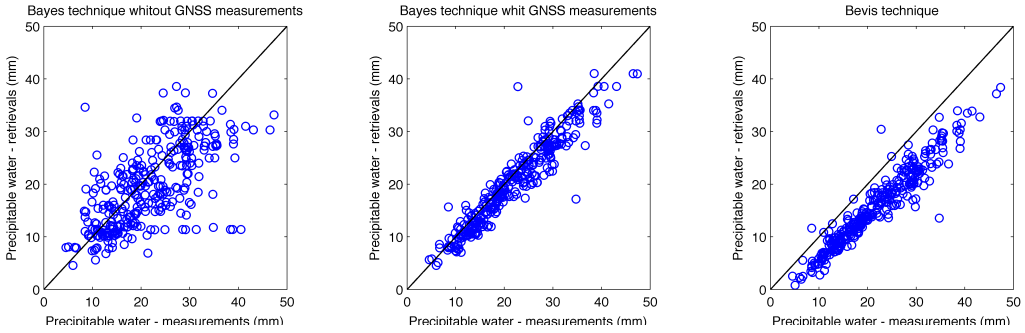


(a) 00:00 UTC

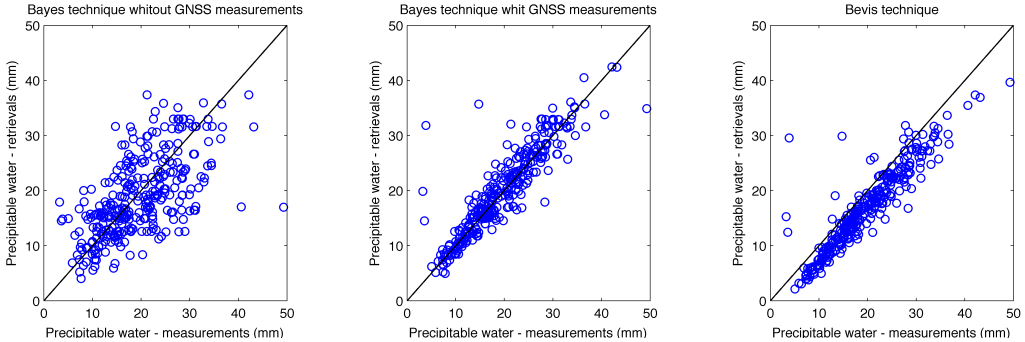


(b) 12:00 UTC

Figure 5.18: Scatter plots of PW best estimates versus balloon measured values of the whole 2011 year using only the basic ground measurement (left panel), including also the GNSS ZPD measurement (middle panel). Corresponding PW estimate obtained using Bevis method are also shown (left panel). First configuration method. Station is cagl (Cagliari).

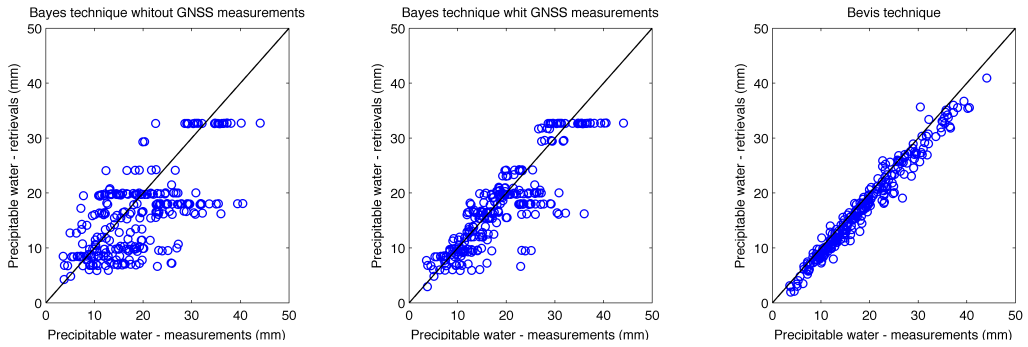


(a) 00:00 UTC

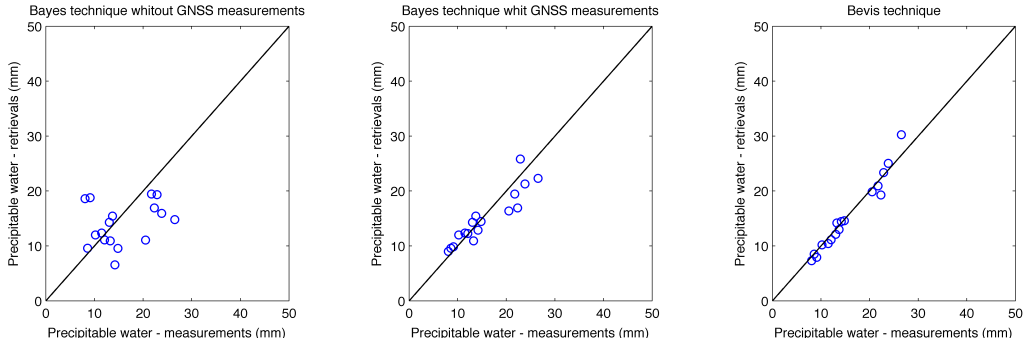


(b) 12:00 UTC

Figure 5.19: Same as Fig. 5.18 but for station of ajac (Ajaccio).

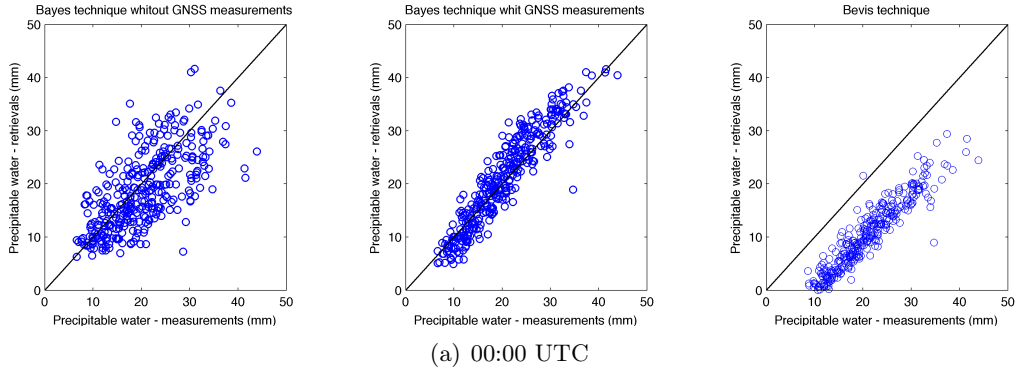


(a) 00:00 UTC

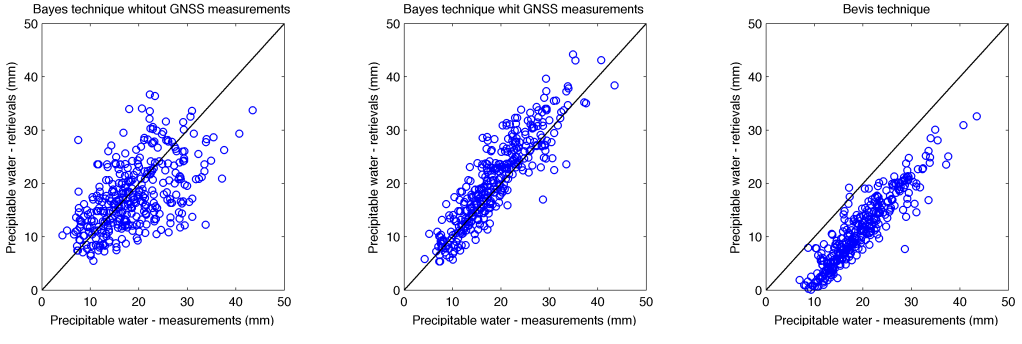


(b) 12:00 UTC

Figure 5.20: Same as Fig. 5.18 but for station of medi (Medicina).

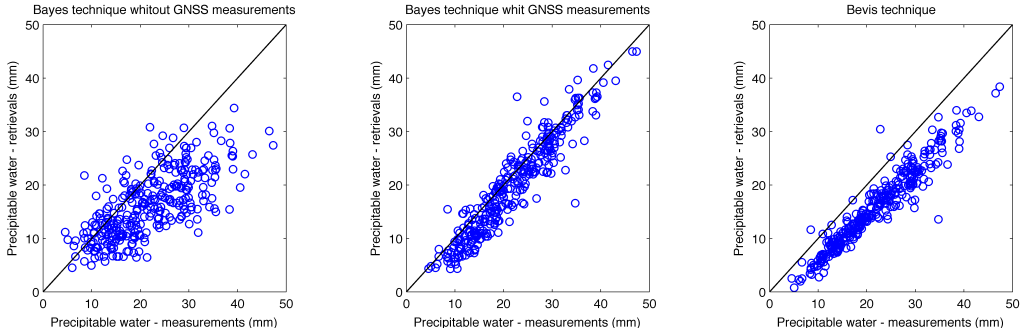


(a) 00:00 UTC

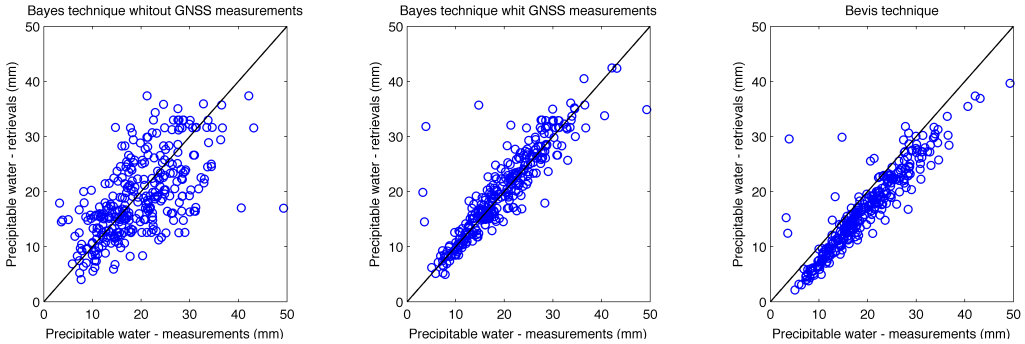


(b) 12:00 UTC

Figure 5.21: Scatter plots of PW best estimates versus balloon measured values of the whole 2011 year both 00:00 (above) and 12:00 (below) time using only the basic ground measurement (left side), including also the GNSS ZPD measurement (middle). Corresponding PW estimate obtained using Bevis method are also shown (left). Second configuration method. Station is cagl (Cagliari).

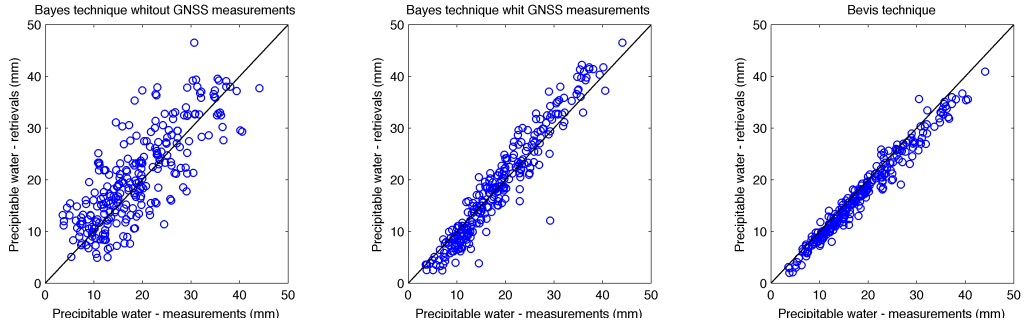


(a) 00:00 UTC

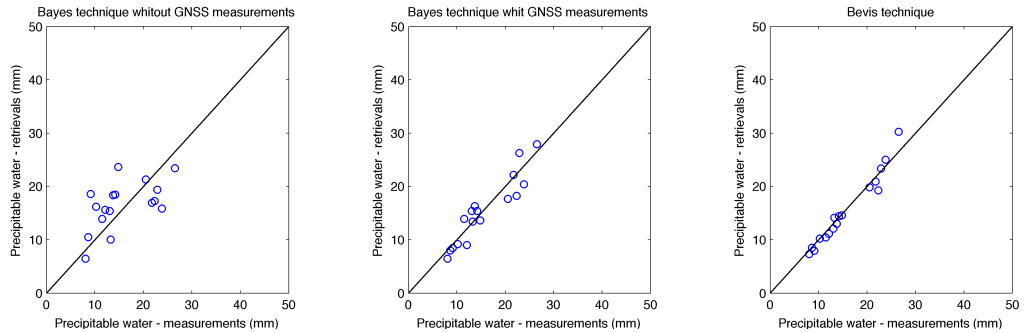


(b) 12:00 UTC

Figure 5.22: Same as Fig. 5.21 but for station of ajac (Ajaccio).



(a) 00:00 UTC



(b) 12:00 UTC

Figure 5.23: Same as Fig. 5.21 but for station of medi (Medicina).

The values obtained by balloon referred as measured values are retrieved from the vertical sounding observation by approximating the integration over the vertical line with a discrete summation of the balloon sampled water density values, multiplied by

the distance between the two subsequent samples.

The introduction of the GNSS measurements improves not only the estimate precision but also its accuracy, making the linear fit very close to the perfect agreement line.

As a further term of comparison, the standard approach ([1]) was applied to the same set of data using, as weighted mean temperature for the atmosphere, the value suggested in [37], related to the ground temperature T_0 through a linear relationship optimised on a dataset of radio-sounding measures over Italy. This method of PW retrieval is described in § 3.3.

A first qualitative evidence of the success of the retrieval emerges fairly immediately by the analysis of the scatter plots. However a quantitative analysis was performed for all stations by calculating some parameters like the correlation (ρ) between the estimated values and the measured ones, showing how the retrieval is able to reconstruct the water vapour content of the atmosphere.

A second important parameter is the Root Mean Square Error obtained as:

$$RMSE_{PW} = \sqrt{\frac{1}{S}\sum_{i=1}^{S}(PW_{ei} - PW_{mi})^2} \tag{5.6}$$

being S the number of retrievals, PW_{ei} and PW_{mi} the estimated (relative to the most probable profile) and the measured values of precipitable water respectively, for the i-th retrieval. Finally the bias has been computed as the quantity given by:

$$BIAS_{PW} = \frac{1}{S}\sum_{i=1}^{S}(PW_{ei} - PW_{mi}) \tag{5.7}$$

The values of these parameters for all the test sites and for both the method configuration are reported in Table. 5.2 and Table. 5.3.

More specifically the root mean square residual decreases and the correlation increases with a strong reduction of the bias, when adding the ZPD measurement to the set of ground-based observations. In the Bevis method the correlation is always higher (in some cases the values are very close to the Bayesian technique ones), but the BIAS and RMSE values are quite high especially for the stations of Ajaccio and Cagliari. The Bevis method in the Medicina station shows a very strong correlation of the retrieved PW values with the measurements, and BIAS and RMSE values lower that in the other test sites. The Medicina GNSS station is sited in an inland area with homogeneous topographical properties, as opposed to Cagliari and Medicina, that are sited close to the coast. The atmospheric conditions are probably more homogeneous and respects the hypothesis according to which the computation of ZPD is made (see § 3.1). In effect the discontinuity surface in a coast line is generally characterised by a discontinuity of the atmospheric water vapour content between land and sea, which can

Station	Variable	Bayes method no GNSS	Bayes method all constrains	Bevis method
Cagliari @ 00	RMSE [mm]	6.5422	2.7666	10.8679
	CORR	0.6169	0.9421	0.9443
	BIAS [mm]	-0.6655	0.8787	-10.5846
Cagliari @ 12	RMSE [mm]	6.3470	3.6012	8.6949
	CORR	0.6071	0.9161	0.9344
	BIAS [mm]	0.1874	1.9840	-8.3074
Ajaccio @ 00	RMSE [mm]	7.0211	3.0366	6.4550
	CORR	0.6549	0.9513	0.9562
	BIAS [mm]	-2.0492	-1.5549	-5.9210
Ajaccio @ 12	RMSE [mm]	6.5808	3.5073	4.1947
	CORR	0.6171	0.8987	0.9190
	BIAS [mm]	-0.5941	0.4148	-2.8731
Medicina @ 00	RMSE [mm]	7.2882	4.5003	3.0960
	CORR	0.6578	0.8706	0.9521
	BIAS [mm]	-2.9018	-1.4840	-1.6143
Medicina @ 12	RMSE [mm]	6.1419	2.4544	1.3751
	CORR	0.3023	0.9197	0.9804
	BIAS [mm]	-1.7397	-0.6533	-0.2225

Table 5.2: RMSE correlation and BIAS values for the PW relative to different stations. Period of analysis is 2011 full year. First test configuration.

affect in a negative way the attempt to reconstruct the ZPD from the measured values of SPD along the various directions, even if in the analysis is introduced the horizontal atmospheric asymmetry (see for instance Eq. 3.3). Note that concerning the Medicina test site, only a few balloon launch data are available at 12:00 UTC for the full 2011 year period. All the available collected data have been used, obtaining a scatter plot made up by only a few values and therefore not so meaningfull.

Station	Variable	Bayes method no GNSS	Bayes method all constrains	Bevis method
Cagliari @ 00	RMSE [mm]	6.2386	3.3165	10.8679
	CORR	0.6634	0.9333	0.9443
	BIAS [mm]	-1.5536	0.6929	-10.5846
Cagliari @ 12	RMSE [mm]	6.5155	4.0609	8.6949
	CORR	0.5479	0.8921	0.9344
	BIAS [mm]	-1.0442	1.7368	-8.3074
Ajaccio @ 00	RMSE [mm]	8.0727	3.5578	6.4550
	CORR	0.7087	0.9436	0.9562
	BIAS [mm]	-5.4362	-1.9107	-5.9210
Ajaccio @ 12	RMSE [mm]	6.6692	4.5058	4.1947
	CORR	0.6632	0.8876	0.9190
	BIAS [mm]	0.493418	1.7185	-2.8731
Medicina @ 00	RMSE [mm]	5.9239	3.0943	3.0960
	CORR [mm]	0.7903	0.9606	0.9521
	BIAS [mm]	0.0546	1.5641	-1.6143
Medicina @ 12	RMSE [mm]	4.9598	2.2320	1.3751
	CORR	0.5721	0.9317	0.9804
	BIAS [mm]	0.8935	-0.2560	-0.2225

Table 5.3: RMSE correlation and BIAS values for the PW relative to different stations. Period of analysis is 2011 full year. Second test configuration.

The second test configuration gives slightly worse performance in all cases with respect to the first configuration test. However, the parameters values and the scatter plots are aligned with those of the first test, giving good results, which are still better than the standard Bevis method in terms of BIAS and RMSE, for the sites of Cagliari and Ajaccio. The advantage with respect to the first configuration is given by the possibility to apply this method to all the point of the grid once the set of measurements

$\tilde{y}$ is given.

In any case the standard method does not provide any probability distribution for the retrieved value of PW, as the developed method does.

Very significant are the results concerning the station of Medicina, for which the estimated values of precipitable water are in some cases very different from the measured reference (see Fig. 5.20), when the first configuration of the method is applied. The poor performance is mainly due to the low number of profiles available to set up the prior dataset and to construct the coefficient array to be used for the retrieval. The dependence of the correlation, RMSE and BIAS parameters from the profile number has been assessed by making a test on the Cagliari station (with a total profile number higher than Ajaccio and Medicina) and artificially varying the number of available profiles (see Fig. 5.24). These artificially limited prior is made with profiles randomly chosen from the whole available prior dataset of the Cagliari site and it constitutes the dataset of profiles to be selected in the retrieval process. The set of coefficients used for this analysis is valid only locally (on Cagliari site), i.e. those coefficient are from the first configuration of the algorithm.

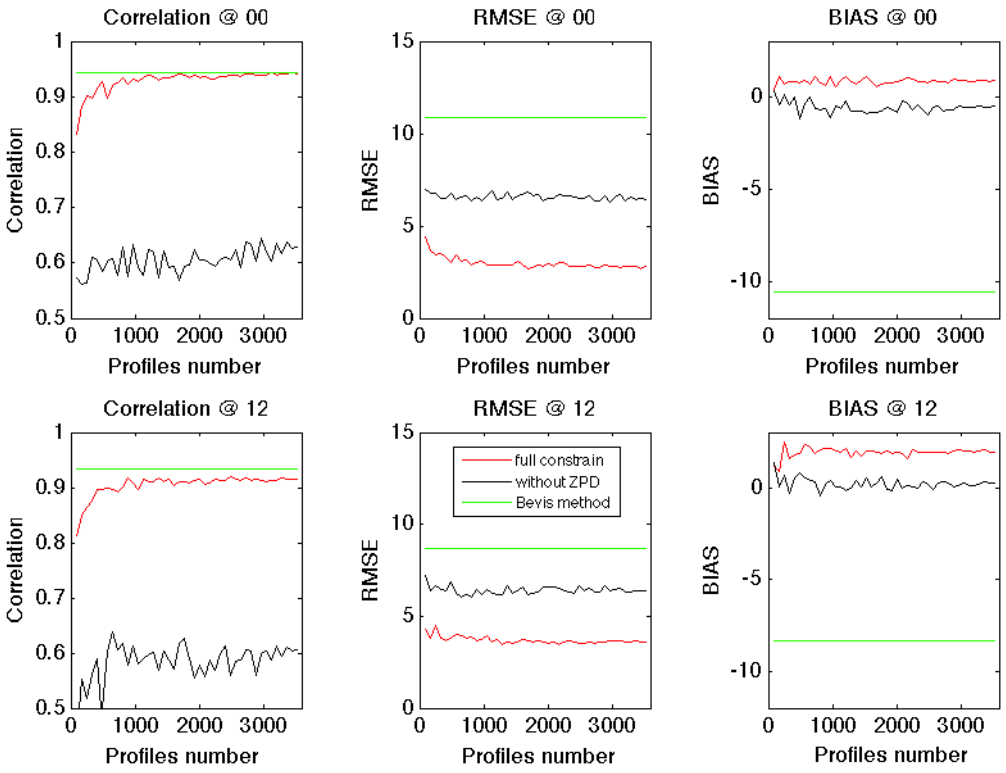


Figure 5.24: PW estimation parameters in function of profile number of the prior dataset. Site of analysis: Cagliari. Period: full year 2011. First test configuration.

Both 00:00 UTC and 12:00 UTC are shown in the figure and the parameters of the retrieval result are represented for the configuration of the only ground-based, the full constraint set of measurement and finally for the standard Bevis method (which is

of course a constant value). The results of this analysis shows an improvement of the performances when the number of profiles increases. This improvement is continuous up to a number of about 1000 profiles for the prior dataset. This can be considered a threshold value of the profile needed to get the best retrieval performances. In fact all parameter values above 1000 profiles are essentially stable, their oscillations become negligible and the gain in terms of accuracy is very little. It is worth to note that the correlation increase up to a threshold of 1000 profiles even in the case in which only the set of ground measurements is used.

5.3 Quantitative validation results

The comparison with balloon observations of the retrieval results for the selected case studies, as shown in § 4.3.3, are useful to identify the qualitative performances of the retrieval. In order to obtain a quantitative assessment of the errors made by the retrieval, a reference measurement is still needed, but also a numeric comparison must be implemented for the error estimation.

For each atmospheric state $\mathbf{x_i}$ of the prior dataset, once a measurement set of the observable vector $\tilde{\mathbf{y}}$ is available, the retrieval determines a certain conditional probability $Pr(\mathbf{x_i}|\tilde{\mathbf{y}})$. Let us call $\mathbf{x_B}$ the corresponding balloon measurements of the state vector coincident in time and in space with the ground-based set of observations. Balloon observations useful for our purposes consist of measurements of pressure, temperature, relative humidity and height of the observation. As the height of the measurements are different from launch to launch and the corresponding pressure level values are generally different from the fixed ones of the MERRA model output, balloon measurements need to be interpolated before being compared with the retrieval results. The so obtained balloon vector is defined as:

$$\mathbf{x_B} = [h_{B1},\ T_{B1},\ e_{wB1},\ h_{B2},\ T_{B2},\ e_{wB2},\ h_{Bj},\ T_{Bj},\ e_{wBj}, ...,\ h_{B_{UP}},\ T_{B_{UP}},\ e_{wB_{UP}}] \tag{5.8}$$

where subscript UP indicates the upper level of the balloon measurement which changes from station to station; h_j is the geopotential height relative to the j-th level and it is computed starting from the quote z_j using Eq. 1.5. The values of e_{wj} have been obtained using Eqs. 1.13 and 1.14 applied to the relative humidity and temperature measurements of the soundings. Note that the j-th component of the $\mathbf{x_B}$ vector is relative to the same isobaric level of the corresponding j-th component of the $\mathbf{x_i}$ vector ($P_{ij} \equiv P_{Bj}$). The state variables for each vertical level are $[T_{ij},\ e_{wij},\ h_{ij}]$ and $[T_{Bj},\ e_{wBj},\ h_{Bj}]$ for the $\mathbf{x_i}$ and $\mathbf{x_B}$ vectors respectively.

The $RMSE$ parameter has been used for the error assessment. As for each state vector a posterior probability is defined and it can be used as weighting function for

the variable V to be analysed (h, T or e_w). The formula for the computation of RMSE is:

$$RMSE_{Vj} = \sqrt{\sum_{i=1}^{M} Pr(\mathbf{x_i}|\tilde{\mathbf{y}}) \left[\mathbf{V_{ij}} - \mathbf{V_{Bj}}\right]^2} \tag{5.9}$$

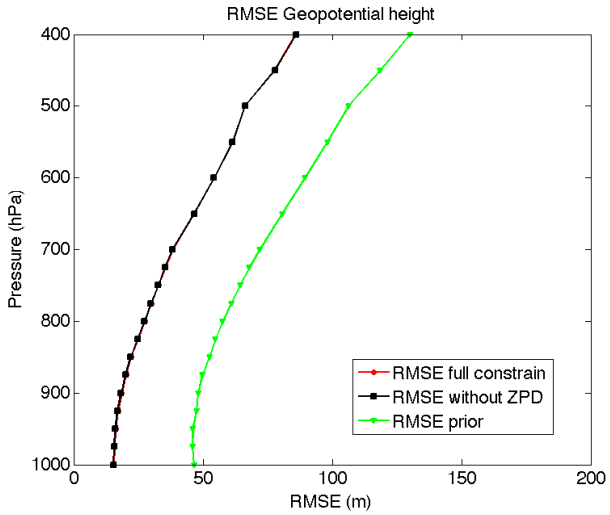
The total $\mathbf{RMSE_V}$ vector is given by all the $RMSE_{Vj}$ components, with j varying from 1-th to UP-th level. These vectors are relative to a specific retrieval results and vary case by case. An average value of $\mathbf{RMSE_V}$ over a time period (e.g. the full 2011 year period) gives a more useful information about the accuracy of the algorithm (see Figs. 5.25 – 5.30).

The plots of the temperatures show a discontinuity in both configurations of the algorithm, at about 200 hPa with a localised $\mathbf{RMSE_T}$ peak (not displayed in the graph for the Ajaccio site whose balloon profile is limited to 400 hPa), in correspondence of the tropopause where the vertical gradient of the temperature changes its sign. For this reason the temperature profile reconstruction errors are higher at this level.

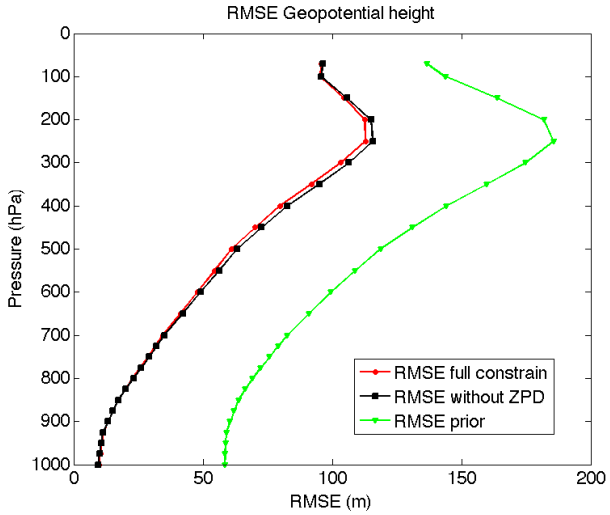
In both cases the temperature profile is poorly reconstructed from the algorithm in the Medicina station, even if some improvements are achieved with the intensification of the prior dataset. In fact, on this site, the dataset of temperature profiles seem to be poorly representative of the possible real profiles. The reconstruction of temperature profiles is better in the other two test sites, where the error is reduced mainly in the levels closest to the surface. This is because the weather station measurements have a strong correlation with the lower atmospheric state values.

Excellent results are obtained for the $\mathbf{RMSE_{e_w}}$, showing coherence when switching from one site to the other. Particularly relevant are the $\mathbf{RMSE_{e_w}}$ results for Medicine, in which the second test gives better performances than the other stations.

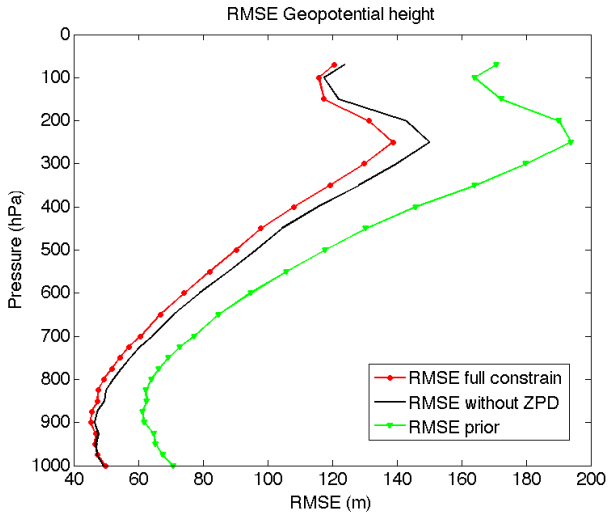
In case of h and T the trend of **RMSE** in function of height remains unchanged including also the GNSS in the retrieval, and in particular it is very similar to that of a priori dataset in which all profiles are equally probable. However the **RMSE** values are reduced also significantly in some cases. The $\mathbf{RMSE_{e_w}}$ have instead different trend in function of height, with respect to the prior dataset. These considerations confirm that the water vapour content in the atmosphere is difficult to model and retrieve without the availability of proper measures. It also confirms that the reconstruction of the profiles of water vapour content is enough accurate to allow to obtain fairly accurate estimates of precipitable water (as shown in § 5.2), which is a particularly relevant result in the perspective of mapping such a significant parameter also in assimilated meteorological forecasts.



(a) ajac

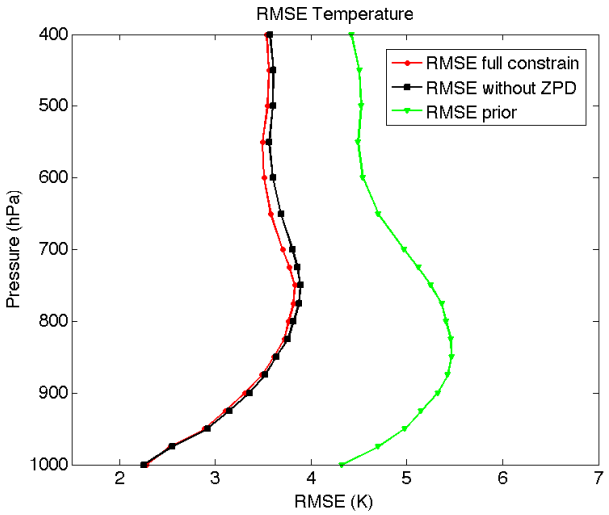


(b) cagl

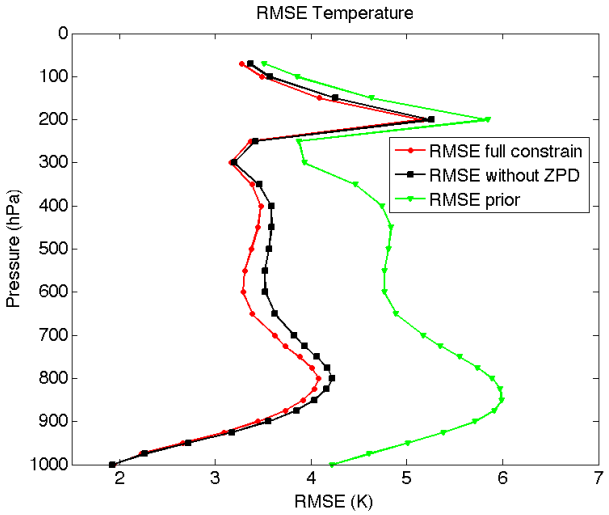


(c) medi

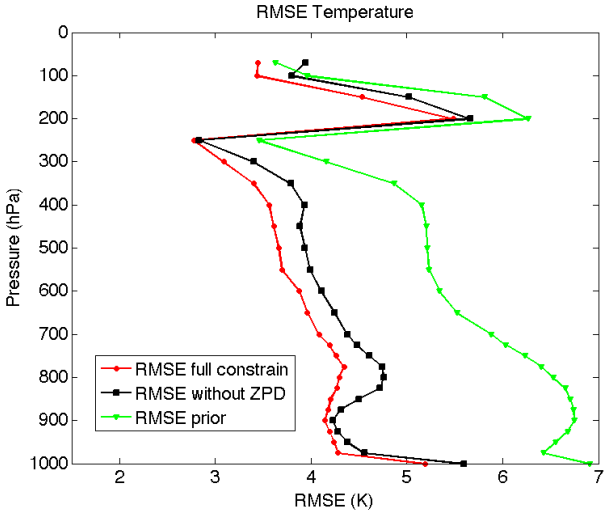
Figure 5.25: Averaged RMSE values computed for the h variable over the full year 2011 including both 00:00 and 12:00 UTC observations. First test configuration.



(a) ajac

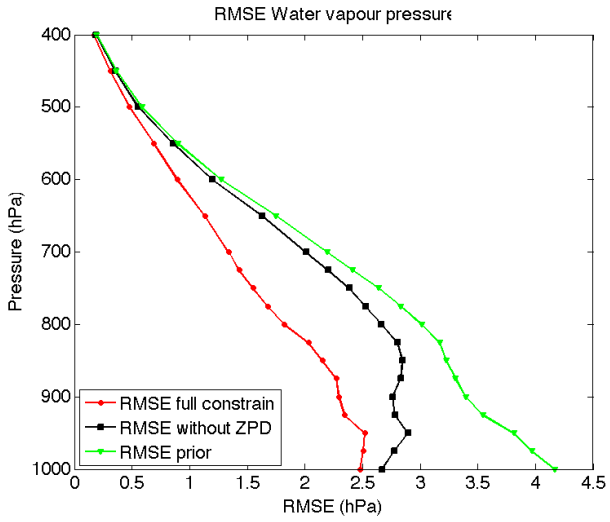


(b) cagl

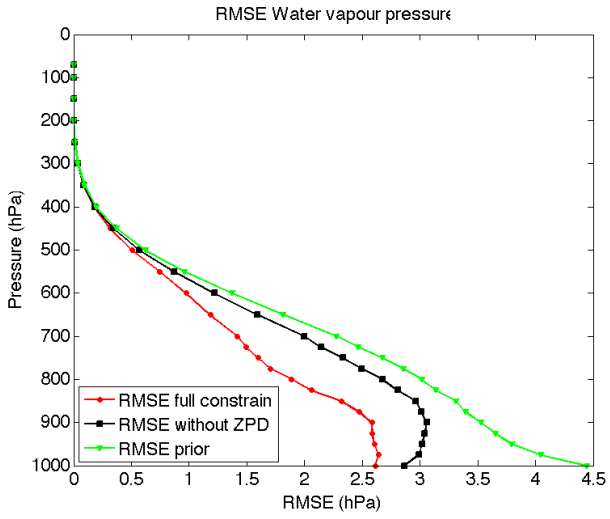


(c) medi

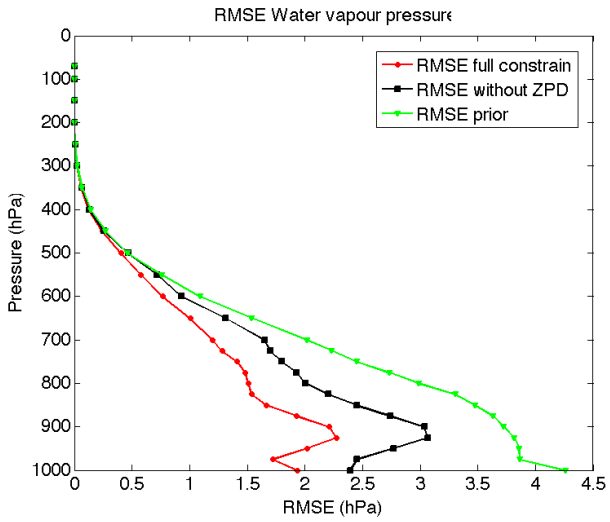
Figure 5.26: Averaged RMSE values computed for the T variable over the full year 2011 including both 00:00 and 12:00 UTC observations. First test configuration.



(a) ajac

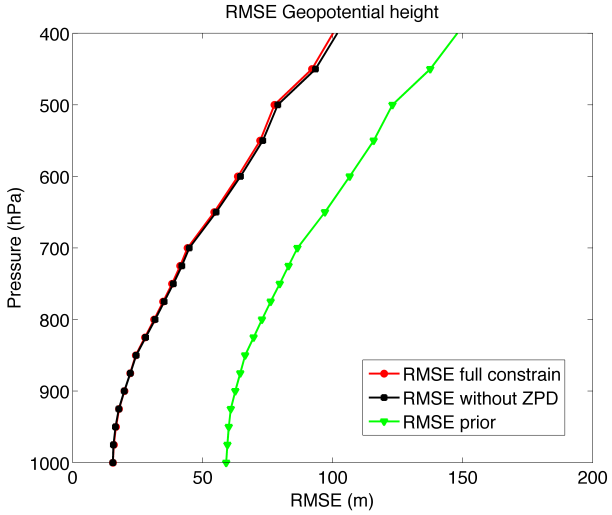


(b) cagl

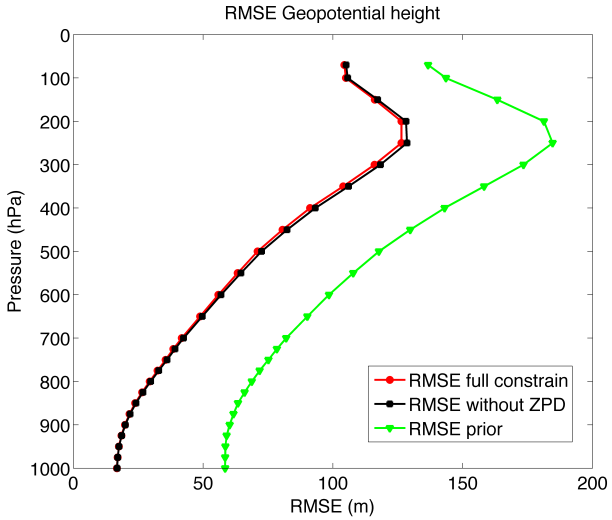


(c) medi

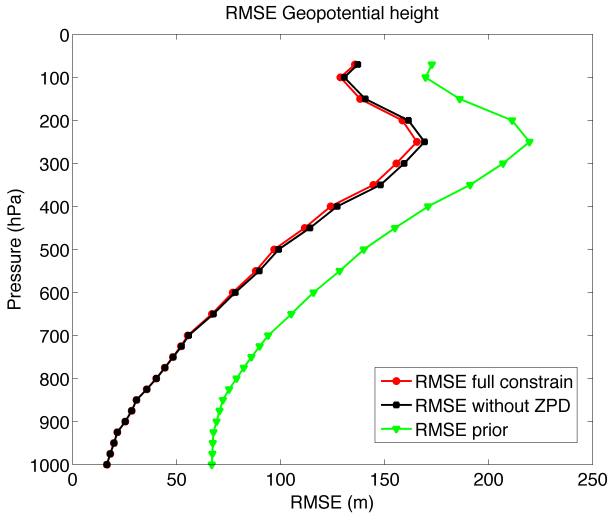
Figure 5.27: Averaged RMSE values computed for the e_w variable over the full year 2011 including both 00:00 and 12:00 UTC observations. First test configuration.



(a) ajac

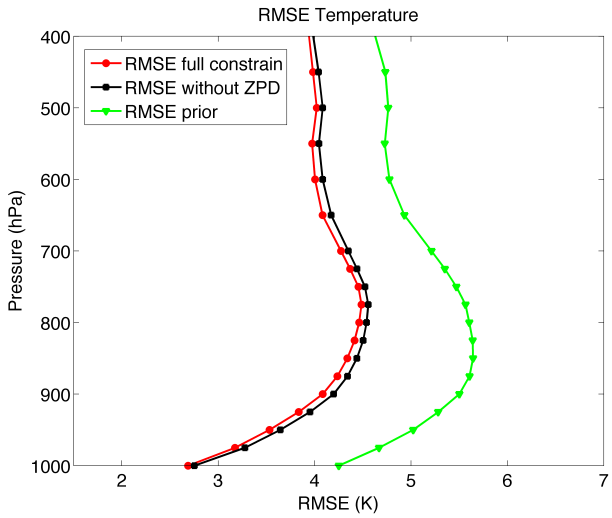


(b) cagl

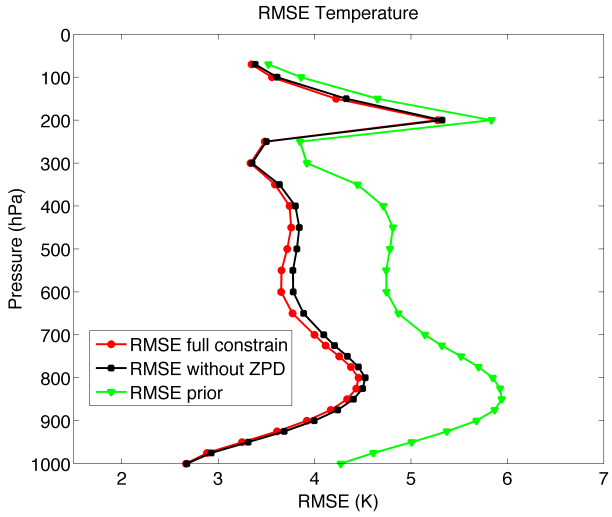


(c) medi

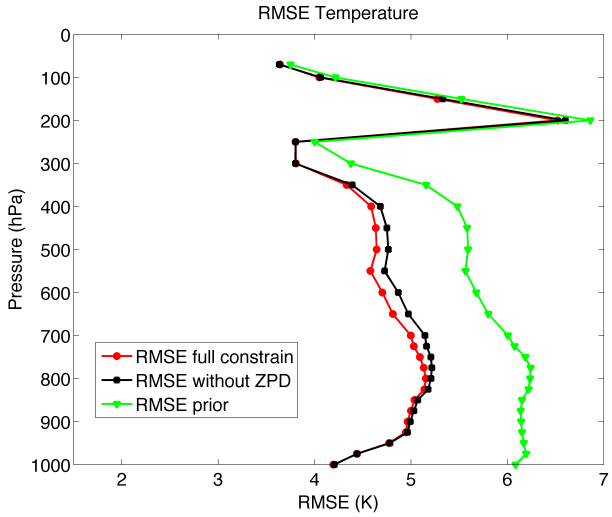
Figure 5.28: Averaged RMSE values computed for the h variable over the full year 2011 including both 00:00 and 12:00 UTC observations. Second test configuration.



(a) ajac

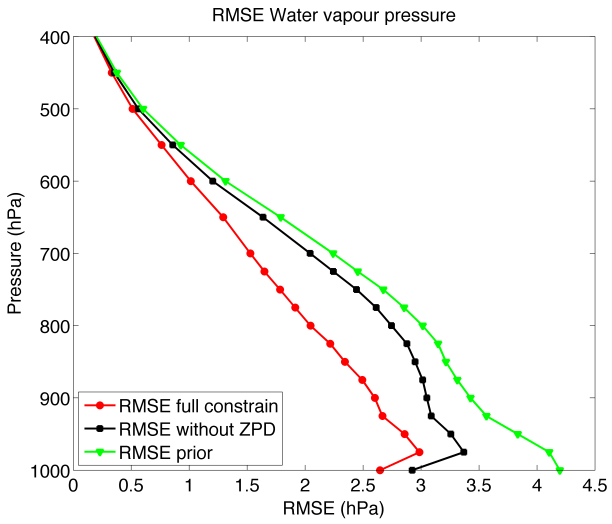


(b) cagl

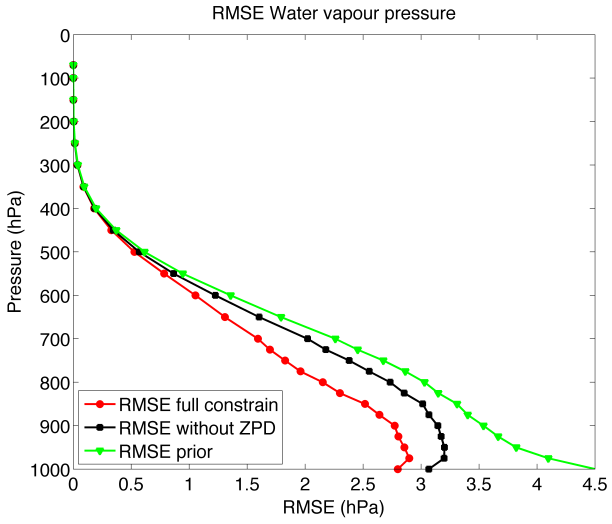


(c) medi

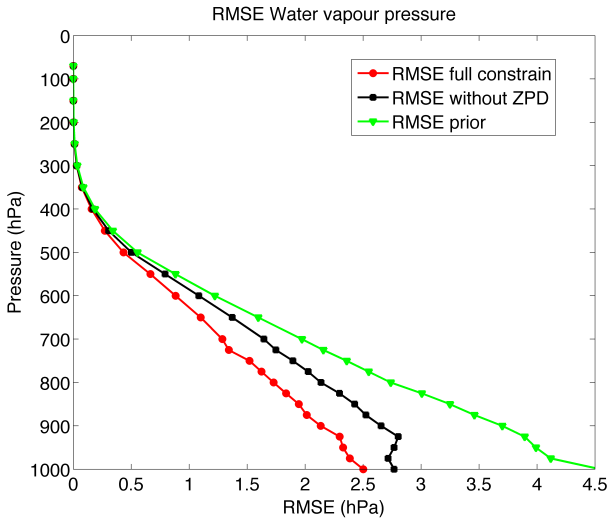
Figure 5.29: Averaged RMSE values computed for the T variable over the full year 2011 including both 00:00 and 12:00 UTC observations. Second test configuration.



(a) ajac



(b) cagl



(c) medi

Figure 5.30: Averaged RMSE values computed for the e_w variable over the full year 2011 including both 00:00 and 12:00 UTC observations. Second test configuration.

Conclusions

In this work we have proposed a novel method for retrieving profiles of water vapour and temperature in troposphere from GNSS data and surface meteorological measures. Namely it exploits the Bayesian inference in order to extract information from ground-based measurements of pressure, temperature, humidity and the total delay of GNSS signals, due to the atmospheric refraction at MW frequencies. The novelty of such a developed method compared to classical approaches are highlighted in the following items:

- **Profiling**: The possibility to provide not only integrated values but also profiles, with a dense sampling.
- **Related errors**: From the retrieved quantities, as they are provided with an associated probability, it is possible to reconstruct the associated errors (and not only a general accuracy estimation through some validation campaigns).
- **Robustness and flexibility**: The method can work with some data missing, without needs of any additional assumption, but with some accuracy degradation, as a consequence.

These properties are particularly relevant for meteorological and climatological applications, e.g. data assimilation into operational forecasts and local climatology assessments.

A decade (2001-2010) of MERRA reanalysis data has been used as a priori knowledge of likely atmospheric states. The results of a first test of operational implementation of the method are shown, applied to local measurements for one year (2011), not belonging to the decade used to build the a priori. Three test sites have been selected (Cagliari, Ajaccio, Medicina) where a meteorological station and a scientific IGS GPS receiver were available together with radiosoundings data to make a comparison of the resulting vertical soundings. Two possible implementations have been tested. The first is only locally applicable, i.e. the retrieval is feasible only in the pixel site used to construct the likelihood function for the bayesian inference. The results of this method, as compared with balloon observations, show high accuracy in the reconstruction of vertical atmospheric variables. The second method is applicable over an extended area, thus

not necessarily only where there are historical data from the measurement stations used for the likelihood function, but everywhere data from a ground meteorological station and a GNSS receiver are contemporarily available. In this second case the accuracy is a bit lower respect to the first case, as expected, because of the use of a series of data coming from different sites, characterised by different climatical and morphological situations.

The method provides a probability distribution for any retrieved quantity, also when the whole set of measurements is not completely available, as mandatory for example for running operational data assimilation into forecast processes. The tests made on the IPW quantity show a very satisfactory agreement between retrieved and measured values, with a significant improvement given by the presence of the GPS delay information, as expected. A valuable agreement is also found in the retrieval of atmospheric profiles of WV and temperature. In a large number of cases the most probable profile actually comes out very close to the measured values, at least up to the tropopause level. A preliminary error evaluation has been made by comparing the retrieved profiles with the real measurements available from balloon soundings, over a full year period (2011). The RMSE, correlation and bias have been used as validation parameters. They show promising performances, mainly in the water vapour vertical content retrieval, but also in the reconstruction of the temperature profile, in favor of the robustness of the method, which gives retrieved quantities that are mutually and physically consistent.

The main drawback of the proposed method, when applied using local data customised for a specific site (first test case), is the necessity of a large dataset for the possible atmospheric states, as a priori knowledge, representing in some sense the local climatology. In this case a proper dataset should be provided for every target point. When there is not enough availability of historical data and the region of interest is large, this drawback becomes a severe limit for the method application. However this limit can be overcome using an extended dataset constructed for a wide area, and extracting a likelihood function using all the available measurements over the available time period. The result is a likelihood probability distribution certainly broader, but with the advantage of being representative of an arbitrarily extended area. This result is much more satisfactory, as more measurement data are used together with the model observations in the construction of the likelihood. In these terms, in the specific case of this work, considerable improvement can be made, as the method as been applied over an area of about 1000 km $\times$ 1000 km, using only 13 IGS station for the likelihood determination. Unfortunately, despite of the availability of a certain number of scientific GNSS receiver stations (e.g the EUREF network, and local geodetic networks) over the selected area, no further ZPD data sources have been found compatible with

the IGS product. In order to exploit this additional amount of information, it should be necessary to reprocess all raw receiver data coming from different networks using an unique software to get a homogeneous product, but this certainly would require the set up of a procedure for ZPD estimation, and consequently a high amount of work. On the other hand a higher number of available ZPD measuring station certainly could help to achieve higher accuracy, for example by dividing the whole area in subareas where the number of observations could be sufficient to get a representative as well as enough selective likelihood.

This method, although in a preliminary version, could be straightfully used for a series of meteorological applications spanning from the assimilation in numerical weather models, to nowcasting and intercalibration purposes. Moreover the potentialities are even greater if considering the number of observations that could be available from a cooperative collection of data coming from sensors all around the globe, in particular if including those on board of ships. In fact the scarcity of observations over sea is one of the major limits of the measures available and commonly used in meteorological applications. This limit may be overcome if all the ship would share their weather ZPD measurements (owning the majority of big vessels both weather stations and GNSS receivers), and consequently, using the method described in this work, also a series of estimations of vertical profiles of pressure, temperature and humidity. Note that the retrieval from slow moving platforms is a very challenging issue, it needs a specific data processing and certainly it should exploit the full precision and redundancy of GNSS signals considering GPS, Galileo and GLONASS. This is one of the challenging purposes of the COSMEMOS Project, a two years research and development project co-funded by the European Commission, 7th Framework Programme, managed by the European GNSS Agency (GSA), whose proposal was born in part as a "side product" of the present work. The project is aimed to improve both safety and efficiency of maritime navigation, and this could be accomplished by improving the knowledge and prediction of the meteo-marine parameters available for the ships.

Although in a fully consistent version, already applicable, the technique needs some improvements and further work, especially on the validation side over different areas around of the world thus over more test sites, using not only balloon observations but also different kind of sources, such as satellites, ground based radiometers, etc...

Bibliography

[1] Bevis, M., S. Businger, T. A. Herring, C. Rocken, R. A. Anthes, and R. H. Ware, 1992: *GPS meteorology: remote sensing of atmospheric water vapor using the global posisioning system.* J. Geoph. Research, 97 (D14), 787-801.

[2] Ortolani, A. 2011: *Probabilistic Tomography of Atmospheric parameters from GNSS data*

[3] Antonini A. , R. Benedetti, A. Ortolani, L. Rovai, and G. Schiavon 2013: *Water Vapor Probabilistic Retrieval using GNSS Signals.* IEEE Transaction On Geoscience and Remote Sensing, ISSN: 0196-2892, doi: 10.1109/TGRS.2013.2256363.

[4] Seeber, Güter - Satellite Geodesy - 2nd completely revised and extended edition - Walter de Gruyter, Berlin, New York 2003.

[5] Wallace, J and P. V. Hobbs, 2006: *Atmospheric Science An Introductory Survey - Second Edition*; University Of Washington, 551 pp.

[6] URL: http://www.kipnews.org/2011/06/12/ionospheric-measurements-with-gps-receivers/

[7] URL: http://iri.gsfc.nasa.gov/

[8] URL: http://www.ips.gov.au/Satellite/2/2

[9] Rabbany A.,E. 2002:*Introduction to GPS The Global Positioning System*, Artech House Mobile Comunication Series, Boston, 194 pp.

[10] Xu G., 2007: *GPS Theory, Algorithms and Applications.* 2nd edition, Berlin Heidelberg 2007, Springer-Verlag, 340 pp.

[11] Davis, J. L., T. A. Herring, I. I. Shapiro, A. E. Rogers, and G. Elgered, 1985: *Geodesy by radio interferometry: Effects of atmospheric modeling errors on estimates of baseline length.* Radio Sci., 20, 1593-1607.

[12] Klobuchar, J.A., 1987: *Ionospheric time-delay algorithm for single-frequency GPS Users.* IEEE Transactions, AES-23(3), 325-331.

[13] Petrie E.J. , M. HernĞndez-Pajares, P. Spalla, P. Moore, M. A. King, 2011: *A Review of Higher Order Ionospheric Refraction Effects on Dual Frequency GPS*, Surveys in Geophysics, 32(3), pp 197-253.

[14] Kleijer F. 2004: *Troposphere Modeling and Filtering for Precise GPS Leveling*; NCG Nederlandse Commissie voor Geodesie Netherlands Geodetic Commission Delft.

[15] Allan, D Statistics of Atomic Frequency Standards, pages 221Ð230. Proceedings of IEEE, Vol. 54, No 2, February 1966.

[16] *IEEE Standard Definitions of Physical Quantities for Fundamental Frequency and Time Metrology — Random Instabilities* **IEEE Std 1139^{TM}-2008**

[17] Kline, P. A., 1997: *Atomic Clock Augmentation For Receivers Using the Global Positioning System.* PhD Thesis, Ohio University - Avionics Engineering Center [http://scholar.lib.vt.edu/theses/available/etd-112516142975720/].

[18] Ashby N., *Relativity in the Global Positioning System* Boulder, 2003, Living Rev. Relativity 6. URL: http://www.livingreviews.org/lrr-2003-1

[19] Prasad, R., and M. Ruggieri, 2005: *Applied Satellite Navigation Using GPS, GALILEO, and Augmentation Systems.* Artech House Mobile Comunication Series, Boston, 290 pp.

[20] Bao-Yen Tsui, J., 2000: *Fundamental of global positioning system receivers: a software approach.* New York, 2000, John Wiley & Sons Inc., 240 pp.

[21] ICD, 1993: *Interface Control Document - Navstar GPS Space Segment / Navigation User Interfaces* (ICD-GPS-200, 1993).

[22] Misra, P., and P. Endge, 2001: *Global Positioning System-Signal, Measurements, and Performance.* Ganga-Jamuna Press, Lincoln, Masssachusetts, 390 pp.

[23] URL://http://igscb.jpl.nasa.gov/components/prods.html

[24] URL: http://www.ngs.noaa.gov/ANTCAL/

[25] URL: http://igscb.jpl.nasa.gov/projects/antenna/index.html

[26] URL: http://ifen.bauv.unibw-muenchen.de/research/signal.htm

[27] Dach, R., U. Hugentobler, P. Fridez, and M. Meindl, 2007: *Bernese GPS Software Version 5.0.* Astronomical Institute, University of Bern.

[28] King, R. W. and Y.Bock, 1999:*Documentation of the GAMIT GPS Analysis Software (version 9.8)*; Unpublished, Massachusetts Institute of Technology, Cambridge, Massachusetts.

[29] Lichten, S.M., Y.E. Bar-sever, E.L. Bertiger, M. Heflin, K.Hurst, R.J. Muellerschoen, S.C. Wu, T.P. Yunck, and J.F. Zumberge 1995: *GIPSY-OASIS II: A high precision GPS Data processing System and general orbit analysis tool*; Technology 2006, NASA Technology Transfer Conference, Chicago, I1., Oct. 24-26.

[30] Marini, J.W. 1972: *Correction of satellite tracking data for an arbitrary tropospheric profile* Radio Sci., 7(2), pp. 223-231

[31] Niell, A. E., 1996: *Global mapping functions for the atmosphere delay at radio wavelengths* Journal of Geophysical Research, 101(B2), pp. 3227-3246.

[32] Boehm, J., A. Niell, P. Tregoing, and H. Schuh, 2006: *Global Mapping Function (GMF): A new empirical mapping function based on numerical weather model data*, Geophysical Research Letters, vol.33

[33] U.S. Standard Atmosphere Supplements, 1966, U.S. Government Printing Office, Washington, D.C., 1966.

[34] Niell, A. E., 2001: *Preliminary evaluation of atmospheric mapping functions based on numerical weather models*, Phys. Chem. Earth, 26, p. 475Ð480.

[35] Boehm, J., B. Werl, and H. Schuh 2006: *Troposphere mapping functions for GPS and very long baseline interferometry from European Centre for Medium-Range Weather Forecasts operational analysis data*, J. Geophys. Res., 111, B02406, doi:10.1029/2005JB003629

[36] Byun S. H. and Y.E. Bar Sever, 2009: *A new type of troposphere zenith path delay product of the International GNSS Service*, J. Geod 83:367-373 DOI 10.1007/s00190-008-0288-8

[37] Basili P., S. Bonafoni, V. Mattioli, P. Ciotti, F.S. Marzano, G. dÕAuria, N. Pierdicca, L. Pulvirenti, 2002: *Assimilazione dati da misure GPS a terra e radiometriche da satellite per la produzione di mappe di vapor d'acqua atmosferico.* Atti della XIV RiNEM, pp. 532-535.

[38] Santos P., K. Carey, W. MacKenzie, J. Zhang, R. Ferraro, J. Yoe, 2007: *Summary of Global Positioning System (GPS) Integrated Precipitable Water (IPW)*, in NWA Electronic Journal.

[39] Hirahara, K. 2000:*Local GPS tropospheric tomography* Earth Planets Space 52(11), 935Ð939.

[40] Flores, A., G. Ruffini, and A. Rius, 2000: *4D tropospheric tomography using GPS slant wet delays.* Annales Geophys- icae 18(2), 223Ð234.

[41] Bender, M., G. Dick, J. Wickert, M. Ramatschi, M. Ge, G. Gendt, M. Rothacher, A. Raabe, and G. Tetzlaff, 2009: *Estimates of the information provided by GPS slant data observed in Germany regarding tomographic applications*, J. Geophys. Res. 114, D06303.

[42] Nilsson, T. and L. Gradinarsky, 2006: *Water Vapor Tomography Using GPS Phase Observations: Simulation Results*, IEEE Trans. Geosci. Remote Sens. 44(10 Part 2), 2927Ð 2941.

[43] "COST Action 716: final report" (URL: http://www.oso.chalmers.se/ kge/cost716.html/)

[44] Kalnay E., Atmospheric modelling, data assimilation and predictability. Cambridge University Press. 2003. pp. 363

[45] Cucurull, L., F. Vandenberghe, D. Barker, E. Vilaclara, A. Rius, 2004: *Three-Dimensional Variational Data Assimilation of Ground-Based GPS ZTD and Meteorological Observations during the 14 December 2001 Storm Event over the Western Mediterranean Sea.* Mon. Wea. Rev., 132, 749Ð763. doi: http://dx.doi.org/10.1175/1520-0493(2004)132<0749:TVDAOG>2.0.CO;2

[46] Lucchesi, R., 2012: *File Specification for MERRA Products.* GMAO Office Note No. 1 (Version 2.3),82 pp, available from http://gmao.gsfc.nasa.gov/pubs/office_notes.

[47] Stockli R. , E. Vermote, N. Saleous, R. Simmon and D. Herring (2005). The Blue Marble Next Generation - A true color Earth dataset including seasonal dynamics from MODIS. Published by the NASA Earth Observatory. http://eoimages.gsfc.nasa.gov/ve//7101/readme.pdf

[48] URL:http://weather.uwyo.edu/upperair/sounding.html

[49] URL: http://www.lamma.rete.toscana.it

[50] URL:http://www.cfr.toscana.it/

[51] URL:http://www.arpa.emr.it/sim/?idrologia/dati_e_grafici

[52] URL:http://www.arpa.emr.it/sim/?osservazioni_e_dati/dati_stazioni_regionali

[53] URL:http://www.dicca.unige.it/meteo/text_files/dati

[54] URL: http://old.politicheagricole.it/ucea/forniture/index3.htm

[55] Thayer, D., 1974: *An improved equation for the radio refractive index of air.* Radio Sci., 9, 803-807.

[56] Andrews D. G., 2010 *An Introduction to Atmospheric Physics - Second Edition*, Cambridge University Press, New York. 249 pp.

[57] URL: http://en.wikipedia.org/wiki/Appleton-Hartree_equation

[58] Bosy J., Rohm W., Sierny J., Kaplon J., 2011: *GNSS Meteorology* TransNav - International Journal on Marine Navigation and Safety of Sea Transportation, Vol. 5, No. 1, pp. 79-83.

[59] Brunner F. K., G. Min, 1991: *An improved model for dual frequency ionospheric correction of GPS observations* Manuscripta Geodaetica Vol.16, pp 205-214.

[60] Davies, K., 1990: *Ionopheric radio. 1. Radio waves. Propagation of effects on ionosphere*, Peter Peregrinus Ltd., London, United Kingdom.

[61] Elgered, G., J. L. Davis, T. A. Herring, and I. I. Shapiro, 1991: *Geodesy by radio interferometry: Water vapor radiometry for estimation the wet delay.* J. Geophys. Res., 96, 6541-6555.

[62] GLONASS ICD, 1998: *GLOBAL NAVIGATION SATELLITE SYSTEM GLONASS Interface Control Document* COORDINATION SCIENTIFIC INFORMATION CENTER, Moscow.

[63] Khinchin, A. I., 1978: *Fondamenti matematici della teoria dell'informazione.* Cremonese Ed., Roma, 129 pp.

[64] Noll, C. 2010: *The Crustal Dynamics Data Information System: A resource to support scientific analysis using space geodesy*, Advances in Space Research. 45, Issue 12, 1421Ð1440, DOI: 10.1016/j.asr.2010.01.018.

[65] Parkinson, B. W., and J. J. Jr. Spilker, 1996: *Global Positioning System: Theory and Applications, Vols.1 and 2.* American Institute of Aeronautics, 370 LâĂŹEnfant Promenade, SW, Washington, DC.

[66] Rüeger J. M. 2002: *Refractive Index Formulae for Electronic Distance Measurement with Radio and Millimetre Waves*, Unisurv Report S-68, School of Surveying and Spatial Information Systems, University of New South Wales, UNSW SYDNEY NSW 2052, Australia, 1-52.

[67] Saastamoinen, J., 1972: *Atmospheric correction for the troposphere and stratosphere in radio ranging of satellites.* in The Use of Artificial Satellites for Geodesy, Geophys. Monogr. Ser., vol. 15, edited by S. W. Henriksen, et al., 247-251.

[68] Smith E. K. and S. Weintraub 1953: *The Constants in the Equation for Atmospheric Refractive Index at Radio Frequencies*, Proceedings of the I. R.E., 41(8), 1035 - 1037 pp.

[69] U.S. Standard Atmosphere, 1976. U.S. Government Printing Office, Washington, D.C.

[70] URL: http://www.esa.int/esaNA/SEM86CSMD6E_galileo_1.html

[71] URL: http://en.wikipedia.org/wiki/GPS

[72] URL: http://tycho.usno.navy.mil/gpstt.html

[73] URL:http://www.navipedia.net/index.php/Principles_of_Interoperability_among_GNSS

[74] Van Der Marel, H., B. Gündlich, 2006:*Slant Delay Retrieval and Multipath Mapping Software* Report for WP6100 - TOUGH - Deliverable D33, Department of Earth Observation and Space Systems (DEOS),Kluyverweg 1, 2629 HS Delft, The Netherlands

[75] Vultaggio, M., 2009: *Lezioni di navigazione satellitare.* Università di Napoli Partenophe - Facoltà di Scienze e Tecnologie, Napoli.