UNIVERSIDAD DE CONCEPCIÓN
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS

ATMOSPHERIC CHARACTERIZATION FOR SITES OF INTEREST IN MILLIMETRIC/SUBMILLIMETRIC ASTRONOMY

Caracterización atmosférica de sitios de interés en astronomía milimétrica/submilimétrica

By: Lenin Andrés Valeria Rivas

Thesis presented to the Facultad de Ciencias Físicas y Matemáticas from the Universidad de Concepción for the degree of Master of Science in Astronomy

September 2023
Concepción, Chile

Advisor: Rodrigo Andrés Reeves Díaz

Abstract

Water vapor is the main source of atmospheric opacity for mm/sub-mm astronomy, hence several studies seek to effectively characterize it for site testing purposes. Reanalysis databases are quickly becoming popular as an alternative to on-site measurements due to easy accessibility and the versatility of the data they provide. In the framework of validating the usage of reanalysis data as a site-testing oriented tool, a statistical comparison of atmospheric water vapor values obtainable from the MERRA-2 database is performed, with ground-based microwave radiometer measurements taken at two astronomical sites in Chile: Llano de Chajnantor, Atacama, and Cerro Paranal, Antofagasta.

MERRA-2 data was interpolated both vertically (across pressure levels) and geographically (latitude-longitude). For each site, different plots are generated: a direct temporal variation plot, to visually compare the data variation over time between both sources; a PWV vs. PWV plot, fitting a linear fit through robust linear regression and calculating both Pearson (r) and Spearman (ρ) correlation coefficients to look for correlations between both data sources; a histogram showing the distribution of the differences between MERRA-2 data and the water vapor measurements (defined as $\Delta PWV = PWV_{\text{MERRA-2}} - PWV_{\text{site}}$), along with its standard deviation (σ), mean (μ), and median values, aiming to better appreciate the similarities of the data sources over time; and a CDF plot to compare both data distributions disregarding timestamps. Finally, mm/sub-mm transmittance curves are created through the *am* atmospheric modeling software which uses ozone and temperature data along with the verified water vapor data for both sites studied, as well as three other sites of interest for the next-generation Event Horizon Telescope: Las Campanas Observatory near La Serena, Chile; Valle Nevado, located near Santiago, Chile; and the General Bernardo O'Higgins base, located in the Antarctica.

The interpolated MERRA-2 PWV values are highly correlated with the ground-based PWV values with a Pearson coefficient greater than 0.9 and a Spearman coefficient higher than 0.9. Their dependence is however, not linear, but $\text{PWV}_{\text{APEX}} = m \cdot \text{PWV}_{\text{MERRA-2}}$, with m being higher than 0.9 in both cases. The difference histograms show an almost zero-centered distribution for Llano de Chajnantor, with a μ value of -0.02 and a median value of -0.007. On the

other hand, in Cerro Paranal, the difference histogram is slightly offset towards positive values, with a μ value of 0.171 and a median value of 0.256. This offset is most likely due to the strong winds present in the site's location, close to the Pacific Ocean. Additionally, the transmittance curves show different performances depending on the site studied. Results obtained at Valle Nevado site suggest promising atmospheric conditions for astronomic observation in the mm/sub-mm range.

The results obtained show that the atmospheric water vapor estimation using MERRA-2 data can be used for site-testing of new sites by evaluating the mm/sub-mm transmittance profile through vertical pressure correction and averaging the closest grid points to the site. This new method opens the door for future site-testing studies using MERRA-2 and potentially other reanalysis databases (i.e. ERA5) as a reliable source of information.

Keywords – site testing, precipitable water vapor, atmospheric model, reanalysis data

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Water vapor is the main source of opacity in the Earth's atmosphere at infrared and mm/sub-mm wavelengths. Hence, the water vapor concentration is commonly considered one of the most relevant selection criteria to evaluate a potential site for a sub-mm telescope, besides the accessibility, radio frequency interference, and wind profile (Li et al., 2020). In general, there is an exponential decrease of the atmospheric water vapor density with altitude up to 2000 m (Starr and White, 1955). Therefore, to obtain very low water vapor concentrations, radio-observatories are commonly installed at high altitudes in very dry zones. However, in contrast to the majority of atmospheric gases, which are well-mixed in the whole atmospheric extension and in near pressure equilibrium, the water vapor concentration varies over time, with altitude, and with geographical location (Kerber et al., 2014). Moreover, the water vapor variability is also directly related to the weather conditions (Smette et al., 2015). Consequently, to determine the atmospheric quality of a site of interest it is required to perform a long-term and continuous site-monitoring of the precipitable water vapor (PWV), which is defined as the depth of the equivalent condensed water (in units of mm) that would result if the total atmospheric water vapor column above the site was condensed.

Well-known areas with PWV conditions suitable for mm/sub-mm astronomy are the Chajnantor area in Northern Chile, Dome A and Dome C in Antarctica, Mauna Kea in Hawaii, Summit in Greenland, and Yangbajing in Tibet (Tremblin et al., 2012). Hitherto, there is a continuous interest in determining new candidates for radio astronomy sites, thus new sites are constantly being evaluated in terms

of PWV, such as the Qitai Telescope (QTT) in China (Li et al., 2020) and the Big Telescope Alt-Azimuthal (BTA) region in Russia (Shikhovtsev et al., 2022). Moreover, there is a high interest in the inclusion of new telescopes in global-scale Very Large Baseline Interferometry (VLBI) networks, such as the Event Horizon Telescope (EHT).

The Event Horizon Telescope (EHT) is a global cooperation network with radio telescopes all over the world that aims to study supermassive black holes (SMBH) located in galaxy centers using Very Long Baseline Interferometry (VLBI). This project has successfully observed M87's central SMBH and Sgr A* in two frequency bands (227.1 and 229.1 GHz). The original EHT was formed by eight observatories at six locations: the Atacama Large Millimeter/submillimeter Array (ALMA) and the Atacama Pathfinder Experiment (APEX) at the Llano de Chajnantor in Chile; the Large Millimeter Telescope Alfonso Serrano (LMT) on Volcán Sierra Negra in Mexico; the James Clerk Maxwell Telescope (JCMT) and Submillimeter Array (SMA) on Mauna Kea in Hawai'i; the IRAM 30 m telescope on Pico Veleta (PV) in Spain; the Submillimeter Telescope (SMT) on Mt. Graham in Arizona; and the South Pole Telescope (SPT) in Antarctica (The Event Horizon Telescope Collaboration, 2019).

Additionally, recent incorporations to the EHT association are Greenland Telescope (GLT) (Matsushita et al., 2017), a 12-meter radio telescope located in the Greenland ice sheet; Kitt Peak Observatory (KPT), a 12-meter telescope located in Arizona, United States; and Northern Extended Millimeter Array (NOEMA), composed of twelve 12-m dishes, located in Plateau de Bure, France (Johnson et al., 2023). Currently, there is an active effort to further expand this network by including $\sim$ 10 new sites and observing new frequency bands (e.g., 86, 230 and 345 GHz) for the so-called next-generation EHT (ng-EHT) (Roelofs et al., 2023), that will serve as an enhanced array that will improve the current EHT temporal sensitivity by $\sim$ 5 orders of magnitude, allowing dynamic analysis with the creation of movies of Sgr A* and M87 through improved "snapshot" imaging, and increasing the sample of black hole shadows imaged.

Important factors to look for in potential sites include the elevation of the main science targets above the horizon, visibility of these targets with the existing sites present in the array and the incremental Fourier coverage contributed by the site, among others. The atmospheric quality of a site is another crucial factor

to considerate when characterizing a new potential site. In this case PWV is the main factor due to its strong presence in the millimeter/submillimeter range. However, since increasing the number of VLBI baselines has more priority than consistent weather conditions, atmospheric performance can be traded off if the site's location is favorable for the project (Raymond et al., 2021).

Traditional site-testing methods to determine the PWV for a specific location rather than an extended geographic area include the use of expensive microwave or infrared radiometers, physically located at the site of interest, that provide continuous measurements. Additionally, another common method is the deployment of radiosondes, that provide complete information of the atmospheric column but lack sufficient temporal resolution. Alternative methods to estimate PWV use the wet delay of the signals received from the global navigation satellite systems (GNSS) (e.g., Özdemir et al. 2018), the GOES satellite data (e.g., Marín et al. 2015), and the MODIS on the NASA Terra and Aqua satellites (e.g., Zhu et al. 2021), among others.

Recent site-testing studies have suggested the use of reanalysis datasets to evaluate the atmospheric water vapor. Raymond et al. (2021) identified a set of possible candidates for the ng-EHT using surface-level PWV historical data from the NASA Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2) (Gelaro et al., 2017). In another recent study, Xu et al. (2022) used measurements from a 183 GHz radiometer along with a site-testing methodology based on the use of reanalysis data obtained from the ECMWF (Li et al., 2020) to evaluate the Muztagh-ata site in China. Iserlohe et al. (2021) compared data obtained from the Far Infrared Field-Imaging Line Spectrometer (FIFI-LS) of the Stratospheric Observatory for Infrared Astronomy (SOFIA), with PWV data extracted from the European Centre for Medium-Range Weather Forecasts (ECMWF). Shikhovtsev et al. (2022) presented a method to evaluate the BTA region using PWV data from the ERA-5 reanalysis dataset. Kuo (2017) used MERRA-2 data to evaluate five sites of interest in mm-wave observation: South Pole, and Dome A in Antarctica, Chajnantor in Atacama, Ali in Tibet and Summit Camp in Greenland. In their study, values for precipitable water vapor, sky brightness temperature, ice water path and liquid water path were calculated to analyze the atmosphere in terms of stability for sky temperature and water vapor content through time. Zhao et al. (2022) used datasets from MERRA-2 and

an atmospheric reanalysis produced by ECMWF's ERA-5 (Hersbach et al., 2020) to evaluate the Lenghu site, a recently discovered site on the Tibetan Plateau.

On the other hand, some studies have used atmospheric radiative transfer models to complement their site-testing evaluation. Matsushita et al. (2017) used MERRA-2 meteorological data for water vapor and ozone mixing, along with the *am* model (Paine, 2022) and measurements of a 225 GHz tipper radiometer to estimate the transmission spectra up to 1.6 THz over the Greenland Summit Camp site of the GLT. Cortés et al. (2016) used MERRA-2 data and the *am* model to simulate atmospheric opacity at 350 μm aiming to correlate opacity measurements from tipper radiometers with atmospheric water vapor in the Chajnantor area, Atacama, Chile.

Additionally, several studies have validated the accuracy of the PWV data obtained from atmospheric reanalyses using ground-based radiometric measurements (e.g., Ssenyunzi et al. 2020; Ermakov et al. 2021). Moreover, Mo et al. (2021) compared 4 reanalysis databases, among them MERRA-2 and ERA5 with GNSS-derived PWV measurements over Antarctica. Indeed, in their study, according to the GNSS-based measurements, MERRA-2 and ERA-5 perform better than the rest of datasets analyzed in terms of PWV estimation.

Nevertheless, although several studies have used PWV data from reanalysis datasets for site-testing purposes, these works are limited by the provided grid and altitude resolution of the dataset. In the present work, a new site-testing method using PWV data from the atmospheric reanalysis dataset MERRA-2 is proposed. Additionally, the PWV data obtained from this methodology will be compared to ground-based 183 GHz radiometer measurements located at two world-class Chilean astronomical sites: APEX, located in Llano de Chajnantor, and LHATPRO, located in Cerro Paranal. The results of this study provide verification of the accuracy of this reanalysis-based site-characterization in terms of water vapor estimation. Moreover, transmittance profiles are generated for the cumulative 50% of cases between 2019 and 2022 over both sites of study, as well as for the three additional sites of interest for the ng-EHT in Chile, previously suggested in near La Serena and Base General Bernardo O'Higgins (GARS) in Antarctica. Indeed, this study is motivated by the determination of new sites for the ng-EHT project, but the methods presented here can be directly applied to the site-testing of any future radio telescope.

1.1 Objectives

The present work has two main objectives:

1. Validation of the usage of reanalysis data for site testing processes in radio astronomy.
   (a) Verify satellite data from the MERRA-2 database using ground-based measurements from two astronomical sites in Chile: Cerro Chajnantor and Cerro Paranal.
   (b) Document and publish the verification process.
2. Design and apply a methodology based on the previously verified atmospheric satellite data, using atmospheric modeling software to obtain transmittance profiles.

Chapter 2

Theoretical background

2.1 Precipitable water vapor (PWV)

Atmospheric water vapor works as a fundamental part of the atmosphere's mm/sub-mm absorption due to the presence of strong emission lines at 22, 183 and 325 GHz, among several others. It's commonly referred to as **precipitable water vapor** (PWV), defined as the total amount of water vapor contained in a vertical column of the atmosphere going from the Earth's surface to the top of the atmosphere (Coakley and Yang, 2014), measured in units of mm.

The water vapor (H_2O) molecule has a bent triatomic configuration with an asymmetric top, as seen in Fig. 2.1.1, in which the oxygen atom is placed in the middle. It has three fundamental vibration modes: symmetric stretching (ν_1), bending (ν_2) and antisymmetric stretching (ν_3). The bending mode is centered at 6.25 μm ($\sim$ 47.97 THz), as seen in Fig. 2.1.2, and is very important in terms of thermal infrared radiative transfer and remote sensing applications. On the other hand, ν_1 and ν_3 fundamentals produce bands centered at 2.74 μm ($\sim$ 109.41 THz) and 2.66 μm ($\sim$ 112.7 THz), respectively. These two bands combine to form a strong band in the solar spectrum, referred to as the 2.7 μm ($\sim$ 111.03 THz) band (Liou, 2002).

Additionally, the water vapor molecule presents three different moments of inertia, given its asymmetric top configuration, leading to three different rotational modes (Fig. 2.1.1). It presents absorption peaks due to resonant interactions, that correspond to this fundamental rotational modes. The lowest spectral line occurs

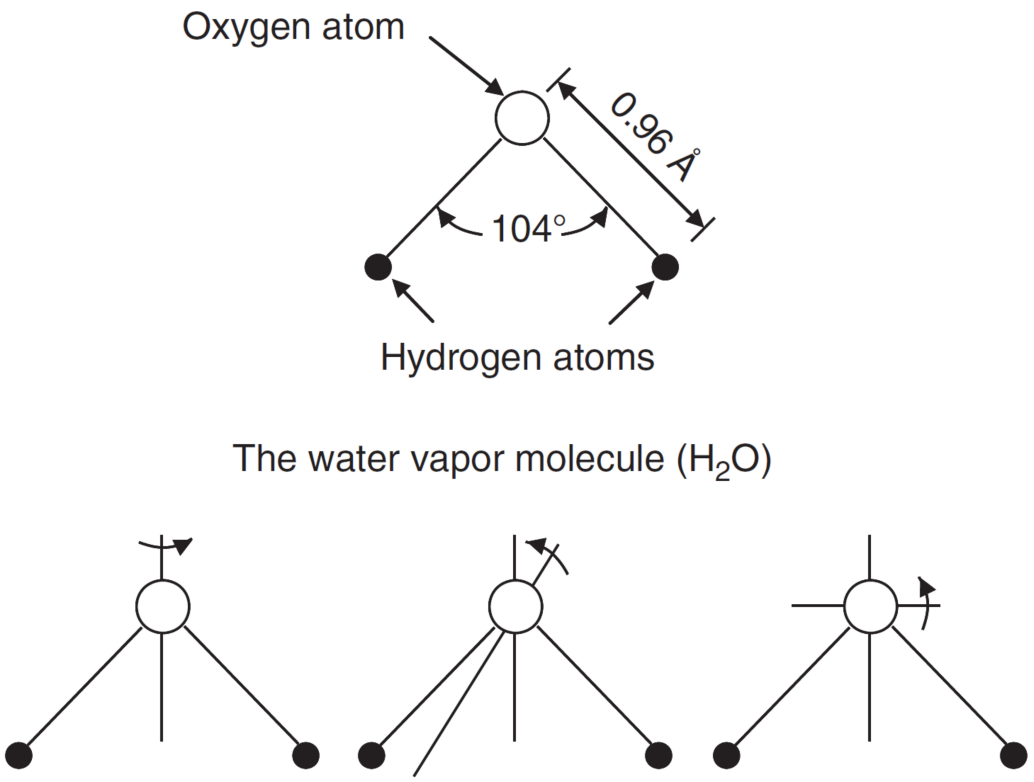


Figure 2.1.1: Extract of an illustrative diagram showing the water vapor molecule and its three rotational modes. (Elachi and Van Zyl, 2021)

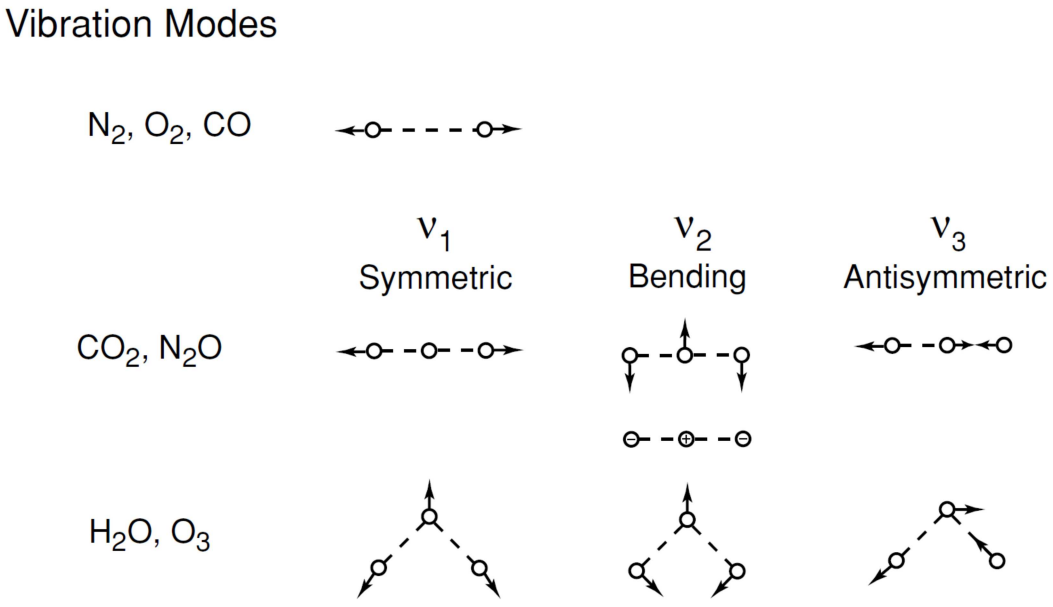


Figure 2.1.2: Extract of an illustrative diagram showing the vibration modes of different atmospheric components, including water vapor. (Liou, 2002)

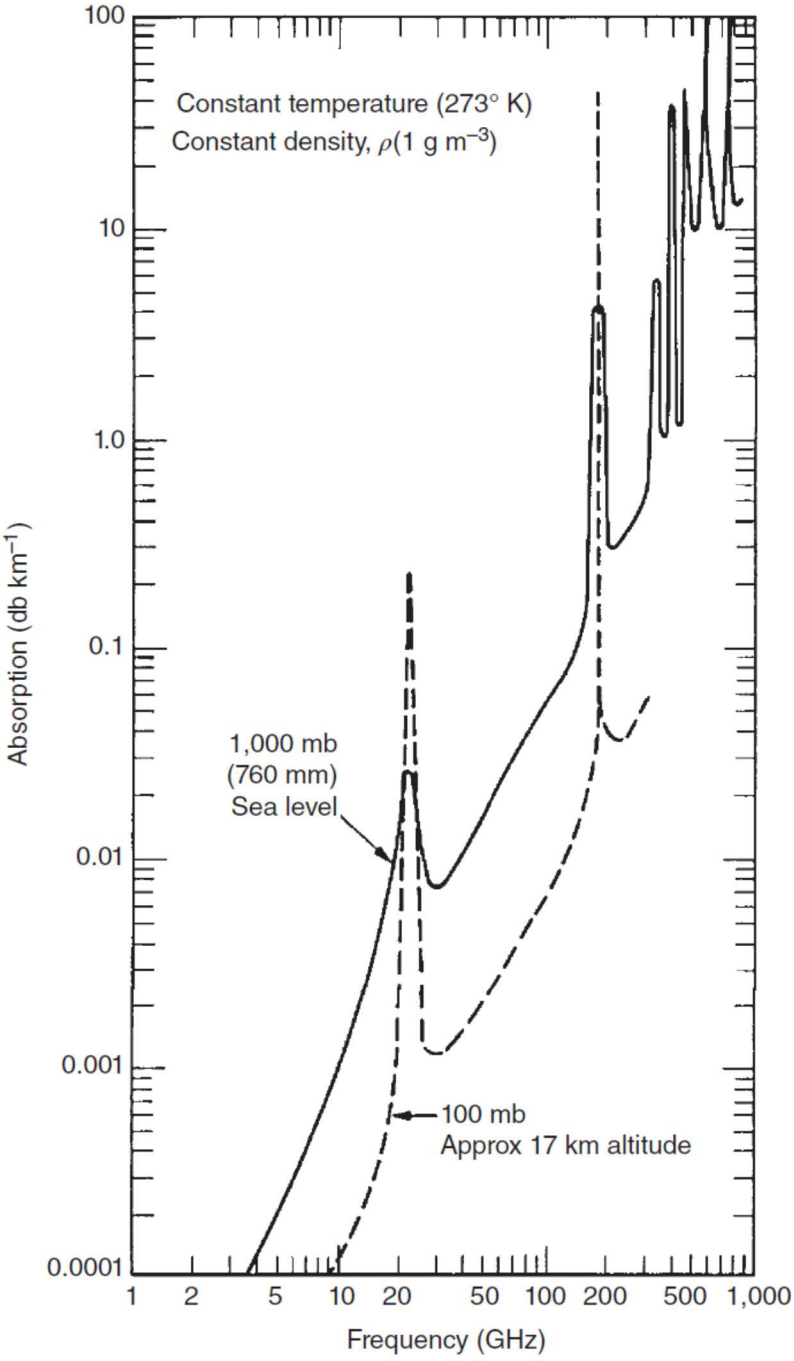


Figure 2.1.3: Absorption spectrum of water vapor at two pressures: 1000 mbar and 100 mbar, at 273 K and for a water vapor density of 1 g/m^3. (Elachi and Van Zyl, 2021)

at ∼22.3 GHz, followed by ∼183.3 GHz and ∼323.3 GHz, with the absorption center increasing the lowest the frequency is, as seen in Fig. 2.1.3. (Otung 2021, Elachi and Van Zyl 2021). This places water vapor as one of the dominant species in terms of absorption in the Earth's lower atmosphere across the microwave spectrum.

For this reason, water vapor is used as the main factor when doing site testing over potential new sites for radio astronomy. PWV has a scale height of ∼ 1 - 2 km (Radford and Holdaway, 1998), meaning that sites above this altitude will present much lower water vapor concentrations than those that are lower. However, water vapor is considered a variable constituent of the atmosphere, strongly depending on weather conditions and geography, thus several studies look to refine its characterization over long periods of time(Kerber et al. 2014, Smette et al. 2015).

For this study, PWV is used as the base of the validation process, comparing atmospheric data from MERRA-2 with real water vapor measurements made in astronomical sites, to analyze the accuracy of this reanalysis data for site testing purposes in radioastronomy. This would put MERRA-2 data as a solid resource for future atmospheric studies regarding site testing and related areas.

2.2 Atmospheric reanalysis

What many see as a common daily weather forecast is actually an initial-value problem that goes through major supercomputers intended for this purpose. Numerical weather prediction (NWP) is a series of processes to predict future meteorological conditions by solving dynamics and physics equations, that provides some guidance for weather forecasting beyond a few hours. NWP is an initial-value problem: given the current state of the atmosphere, a mathematical model simulates and thus predicts its evolution. The determination of these initial conditions is complex and as has become a study field on itself, and has given birth to several different methods used to optimize this process. (Kalnay, 2003)

In early experiments, it became evident that "hand" interpolations with available observations were not efficient due to the great amount of data that had to be manually digitized, that also was very nonuniformly distributed in space (Fig. 2.2.1). Aside from this, there was not enough data to initialize any models. As

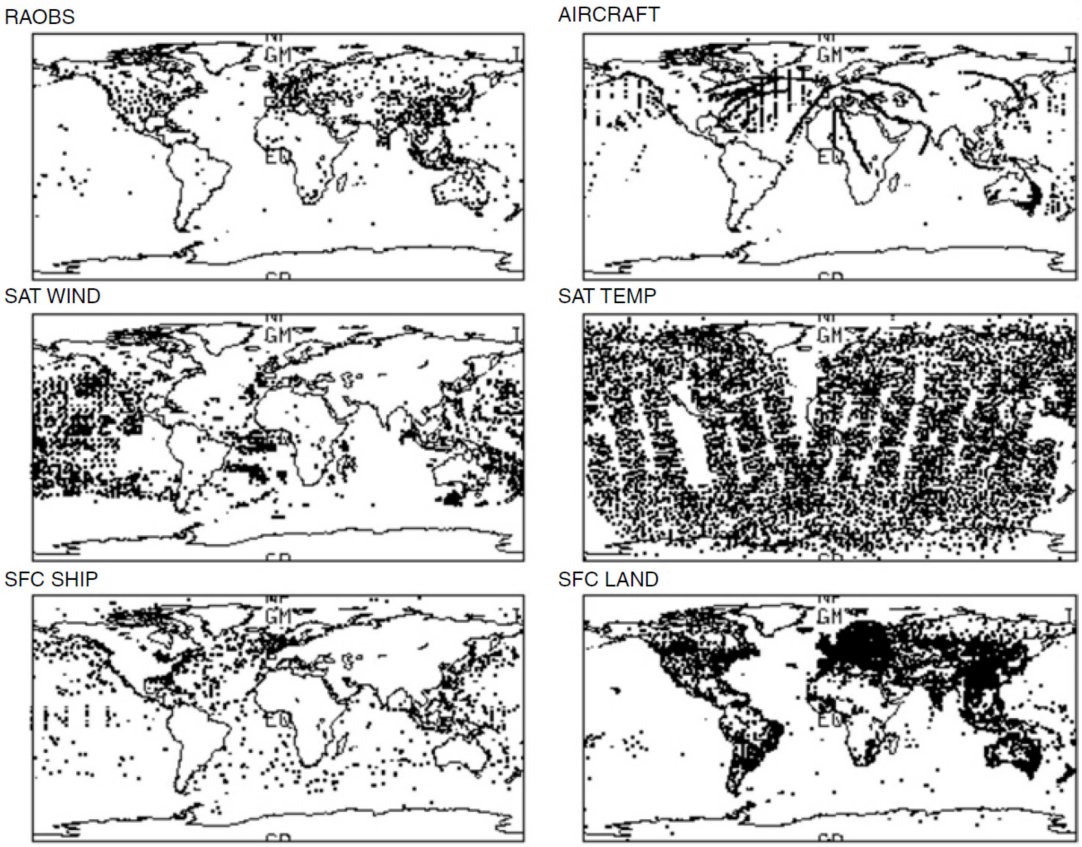


Figure 2.2.1: Typical distribution of observations in a ± 3-hour window. (Kalnay, 2003)

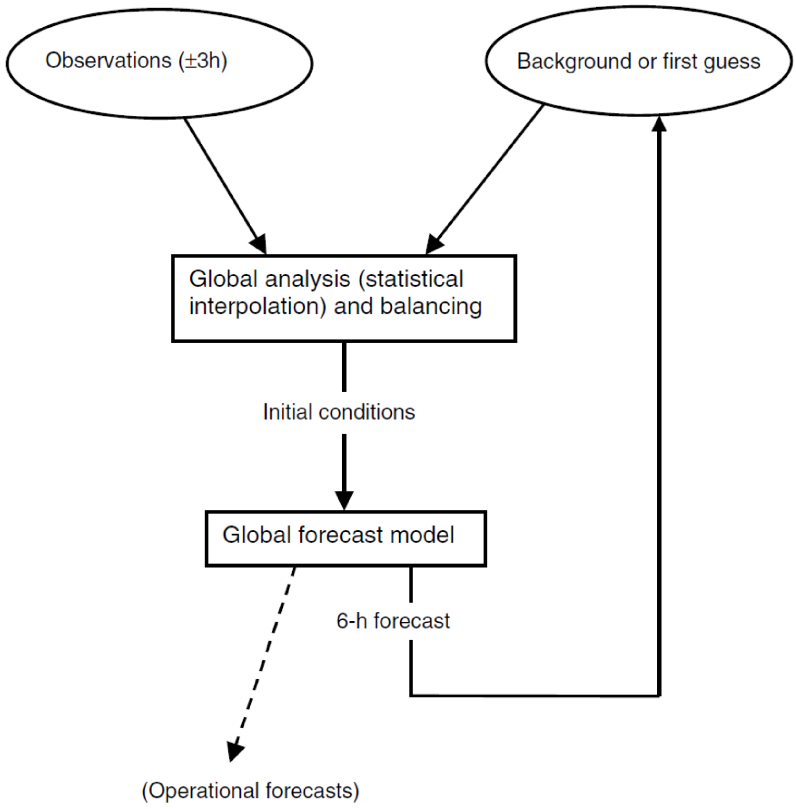


Figure 2.2.2: Flow diagram of a typical intermittent (6-h) data assimilation cycle. (Kalnay, 2003)

an example, for a latitude-longitude model with a resolution of 1° ans 20 vertical levels, there would be $1.3 times 10^6$ grid points. Every grid point has to carry at least 4 prognostic values (two wind components, temperature and moisture), and that would give one 5 million variables that need to be given an initial value. This led to the need for additional information to prepare initial conditions, denoted as *background*, *first guess* or *prior information*. Short-range forecasts were chosen as the first guess in operational data assimilation systems. The intermittent data assimilation cycle, shown in Fig. 2.2.2, is continued in present-day operation systems, that typically use a 6-h cycle

In this 6-h cycle for a global model, the model forecast x^b (tridimensional) is interpolated as a first guess to the observations location, and if they are different, converted from model variables to observed variables y_0. The first guess of the observations is therefore $H(x^b)$, where H is the observation operator that performs the interpolation and transformation from model variables to observation space. The analysis x^a is obtained by adding the "innovations", defined as $y^0 - H(x^b)$, with weights W, determined based on the estimated statistical error covariances of the forecast and observations:

$$x^a = x^b + W[y^0 - H(x^b)] \tag{2.2.1}$$

Different analysis schemes (SCM, OI, 3D-Var and KF) are based on what's seen in Fig. 2.2.2 but differ by the approach taken to combine the background and the observations to produce the analysis Kalnay (2003). To this day, reanalysis has become a staple for the atmospheric research community, not only for climate monitoring, but also for business like energy and agriculture. Reanalyses like the ones from the National Oceanic and Atmospheric Administration/National Centers for Environmental Prediction (NOAA/NCEP), the European Centre for Medium/Range Weather Forecasts (ECMWF) and the National Aeronautics and Space Administration/Global Modeling and Assimilation Office (NASA/GMAO) provide vast ensembles of climate data products. (Gelaro et al., 2017).

2.2.1 MERRA-2

The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2) (Gelaro et al., 2017) is the latest atmospheric reanalysis produced

by NASA's Global Modeling and Assimilation Office (GMAO). It combines the observations made by the GMAO satellite era with the Goddard Earth Observing System (GEOS) atmospheric data assimilation system.

MERRA-2 stores data in a 0.625° longitude x 0.5° latitude horizontal grid that extends across the globe, consisting of 580 values in the longitudinal direction and 360 values in the latitudinal direction. There are four different formats of vertical grid configuration: horizontal-only, which does not provide vertical information; pressure-grid, interpolating data in 42 fixed pressure values; model-level, interpolating data in 72 surface-dependent levels; and model-edge, using the pressure thickness (DELP) to calculate the upper edges of the layers. Horizontal-only gives variables as 3-dimensional fields (x,y,time), while the other configurations give variables as 4-dimensional fields (x,y,z,time). The grid (horizontal and vertical) configuration is defined when retrieving data files from the GES DISC page[1].

Data is distributed over a series of file collections storing different types of data according to the researcher's objective. Each file has a unique index that specifies in detail the data contained and its respective attributes. In this study, the format M2T3NVASM (Global Modeling And Assimilation Office, 2015) was selected, providing 3-hour time averaged assimilated meteorological fields over the model-level vertical grid, in particular specific humidity (QV), that can be converted to PWV along with pressure thickness (DELP), variable that is also present in this format.

A more in-depth description of the contents of this file format as well as other formats can be found in the MERRA-2 file specification document (Bosilovich et al., 2016).

2.3 Radiative transfer

The atmospheric interactions with electromagnetic waves are governed by the propagating wave's wavelength, the physical characteristics of the atmosphere (pressure, temperature, thickness etc.) and its components. These interactions are varied and complex to model given the three-dimensional nature of the atmosphere and the diversity of the interaction mechanisms: scattering, absorption, emission and refraction.

[1] https://disc.gsfc.nasa.gov

The radiative transfer equation is a fundamental equation that describes the propagation of electromagnetic radiation through a **scattering** and **absorbing** medium. At a given point, the change of intensity $I(z, \theta, \phi)$ as a wave travels a distance dz in the direction (θ, ϕ) consist of the following elements:

$$\frac{dI}{dz} = -\alpha_a I - \alpha_s I + \alpha_a B + \alpha_s J \tag{2.3.1}$$

where α_a corresponds to the sum of the absorption coefficient of all the medium components, α_s is the scattering factor of the particles in the medium, $\alpha_a B$ is the portion of energy that's added to the wave due to thermal emission and $\alpha_s J$ is the portion of energy that's added to the wave as a result of scattering of waves from other directions (Elachi and Van Zyl, 2021).

As stated before, the solution of this radiative transfer equation is complex to calculate and often implies the usage of numerical methods and simplifying assumptions. The equation can be simplified for more specific cases by assuming lesser interactions with the incident wave (e.g neglecting scattering). For example, in the case of microwave interactions with the atmosphere, a non-scattering thermal medium can be established, rewriting the last equation in a shorter, simplified way. The variation of the specific intensity I_ν, in this case at the frequency interval for microwaves, at a point s, the radiative transfer equation can be rewritten as:

$$\frac{dI_\nu}{ds} = -I_\nu \alpha + S \tag{2.3.2}$$

where α is an absorption coefficient, describing the loss of energy, and S is a source term, describing the gain of energy into the propagation direction. In the general theory, scattering would grant additional values for gains and losses to the intensity and could be taken into account in the terms of S and α, respectively. However, for calculation purposes, scattering will be neglected in this case. This would consequently turn the source term S into only the locally generated contribution to the radiation and the absorption coefficient α into a scalar characteristic that describes a true loss of energy from the radiation field into the medium (Janssen, 1993). Assuming thermodynamic equilibrium so that each point can be characterized by a temperature T, for balance between emitted and absorbed

energy to be maintained, the S term is strictly required to be

$$S = \alpha B_\nu(T) \tag{2.3.3}$$

where $B_\nu(T)$ is the Planck function:

$$B_\nu(T) = \frac{2h\nu^3}{c^2}\frac{1}{e^{h\nu/kT} - 1} \tag{2.3.4}$$

h in Planck's constant, k is Boltzmann's constant, c is the speed of light, and ν is the frequency. In this case, B_ν and I_ν has the same units. Neglecting the effects of scattering in α and depending only on the intensity along the path of propagation, the solution of this radiative transfer equation can be reduced to:

$$I_\nu(0) = I_\nu(s_0)e^{-\tau(s_0)} + \int_0^{s_0} B_\nu(T)e^{-\tau(s)}\alpha ds \tag{2.3.5}$$

where τ is the optical depth and is defined as:

$$\tau(s) = \int_0^{s_0} \alpha(s')ds' \tag{2.3.6}$$

For this study, a similar case for the radiative transfer equation is used, simulating a plane-parallel non-scattering atmosphere through the radiative transfer model called *am*.

2.3.1 The *am* atmospheric model

The *am* atmospheric model (Paine, 2022) is an open-access and open-source software developed by Scott Paine at the Smithsonian Center of Astrophysics at Harvard University, Cambridge, USA, that will be used in this study to generate the transmittance profiles after the validation process is made. This model solves radiative transfer computations that involve propagation through the atmosphere at microwave wavelengths.

The simplest *am* propagation path is used in this study and is shown in Fig. 2.3.1, also showing the basic spectra that can be computed. Radiation, having a Planck spectrum with a background temperature T_0, propagates through a series of *layers*. These layers represent a segment of a propagation path, mostly associated with atmospheric horizontal strata. Each layer contain a mixture of absorbing species

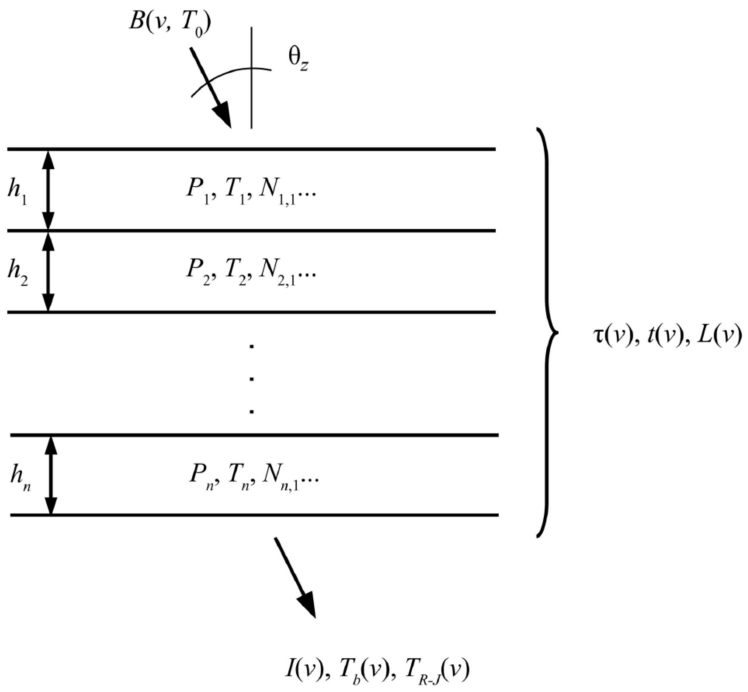


Figure 2.3.1: Illustration showing the *am* layer model.

inside its boundaries. Each type and density of the various elements contained can be individually defined, with each individual definition called a *column*.

In Fig. 2.3.1, each nth layer has its pressure P_n, temperature T_n, column density $N_{n,1}$ and height h_n defined. In turn, this characteristics define the layer stack's opacity $\tau(\nu)$, transmittance $t(\nu)$ and excess propagation delay $L(\nu)$. Once the incident radiation $B(\nu, T_0)$ passes through the layer stack with a θ_z angle, the outputs will be radiance $I(\nu)$, Planck brightness temperature $T_b(\nu)$ and Rayleigh-Jeans brightness temperature $T_{R-J}(\nu)$.

This radiative transfer model in *am* makes several approximations. Local thermodynamic equilibrium is assumed, multiple scattering is neglected and radiation is assumed to be unpolarized. Each layer defines its own opacity or optical depth $\tau_i(\nu)$ as the sum of the opacities of all j columns defined on the layer.

$$\tau_i(\nu) = \sum_j \tau_{i,j}(\nu) \tag{2.3.7}$$

Temperature and pressure can be defined in two ways: At the midpoint of the layer or at the layer boundaries. For this study, temperature and pressure are

defined at layer boundaries as T_{base} and P_{base} respectively, and the Planck function

$$B(\nu, T) = \frac{2h\nu^3}{c^2}\left[\frac{1}{e^{h\nu/kT} - 1}\right] \tag{2.3.8}$$

is assumed to vary linearly with optical depth across each layer. For each layer, the radiance $I_i(\nu)$ at t he lower boundary of the ith layer is:

$$I_i(\nu) = B(\nu, T_{\text{base},i}) \cdot p(\tau_i(\nu)) + B(\nu, T_{\text{base},i-1}) \cdot q(\tau_i(\nu)) + I_{i-1}(\nu) \cdot e^{-\tau_i(\nu)} \tag{2.3.9}$$

where

$$p(\tau) = 1 - \frac{1 - e^{-\tau}}{\tau} \quad , \quad q(\tau) = \frac{1 - e^{-\tau}}{\tau} - e^{-\tau} \tag{2.3.10}$$

The optically thick limit of these expressions is

$$\begin{matrix} p(\tau) \to 1 \\ q(\tau) \to 0 \end{matrix} \quad \text{as } \tau \to \infty. \tag{2.3.11}$$

This radiative transfer computation is done iteratively starting at the top of the layer stack. The opacity spectrum $\tau_i(\nu)$ for each layer is computed, and the initial radiance spectrum is takes as

$$I_0(\nu) = B(\nu, T_0) \tag{2.3.12}$$

Then, Eq. 2.3.9 is applied sequentially through the layers until the spectral radiance emerging from the layer stack is computed.

$$I(\nu) = I_n(\nu) \tag{2.3.13}$$

In addition, the accumulated total optical depth or opacity through the layer stack is computed.

$$\tau(\nu) = \sum_i \tau_i(\nu) \tag{2.3.14}$$

The radiance $I(\nu)$ and opacity $\tau(\nu)$ are the two basic spectra produced by *am*, and a number of other spectra can be derived from them, like an equivalent brightness temperature $T_b(\nu)$ for the radiance $I(\nu)$ inverting Eq. 2.3.8 and substituting $I(\nu)$ for $B(\nu, T)$:

$$T_b(\nu) = \frac{h\nu}{k \ln\left(1 + \frac{2h\nu^3}{c^2 I(\nu)}\right)} \tag{2.3.15}$$

Also, taking the limit of Eq. 2.3.8 when $h\nu \gg kT$, the Planck radiance in the Rayleigh-Jeans is obtained, from which a spectral temperature can be defined:

$$T_{R-J}(\nu) = \frac{c^2}{2k\nu^2} \cdot I(\nu) \tag{2.3.16}$$

On the other hand and directly related to the objective of this study, the transmittance $t(\nu)$ can be obtained based on the opacity or optical depth $\tau(\nu)$ computed. The transmittance is defined as

$$t(\nu) = e^{-\tau(\nu)} \tag{2.3.17}$$

Transmittance is the value that will be used to evaluate the atmospheric performance of different sites when the validation process usin water vapor measurements is completed.

Chapter 3

Methodology

For this study, MERRA-2 data will be cross-matched with real water vapor measurements, looking to validate its accuracy for site testing applications. The sites chosen as data sources for water vapor measurements are Llano de Chajnantor in Atacama, Chile, which houses the Atacama Pathfinder Experiment (APEX) telescope; and Cerro Paranal near Antofagasta, Chile, which houses the Low Humidity and Temperature Profiling (LHATPRO) radiometer. Both sites are of astronomical importance due to them presenting observatories like ALMA (in Llano de Chajnantor's case) and the ESO's Paranal Observatory (in Cerro Paranal's case), while also having water vapor data available in real time and publicly accessible. This validation process consist in the creation of 4 plots to compare the data from both sources in different matters and determine statistical metrics and correlation coefficients for a quantitative analysis.

After the validation process is done, a secondary atmospheric performance evaluation process is made, using previously validated MERRA-2 water vapor data along with the *am* software to generate transmittance profiles for both previously studied sites alongside three additional sites of interest for the ng-EHT, 2 located in Chile and 1 located in Antarctica.

3.1 Data preprocessing

3.1.1 Site measurements: averaging and time matching

Both sites have water vapor measurements with fine time resolution, between ∼1 to 5 seconds between measurements. The chosen MERRA-2 file format provides **time-averaged** data, representing the overall behaviour of the atmosphere within a 3 hour time frame and timestamped in the central time of the interval (Bosilovich et al., 2016). In contrast, the obtained measurement data is **instantaneous**, only showing the water vapor concentration in the exact time it was detected. Thus, to correctly compare both data sources, an averaging process for the measurement data is necessary.

In this study, the data analyzed at both sites corresponds to a period of 4 years, from 2019-01-01 to 2022-12-31. It is relevant to highlight that during this period only weak-moderate el Niño climatic events have been observed. Additionally, a set of incorrectly calibrated PWV data from APEX was not included in the analysis process and few negative PWV values were filtered from LHATPRO. The deleted period from APEX ranges from 14-11-2018 to 06-04-2019, and it was determined through an intercalibration study of several radiometers located at the Chajnantor area (in preparation).

3.1.2 Reanalysis data: conversion and interpolation

Both APEX and LHATPRO provide integrated precipitable water vapor (PWV) values through time. While MERRA-2 also has the option to deliver integrated PWV values over its grid points (in the horizontal-only format used in Raymond et al. 2021), the grid resolution used in its calculus makes the usage these values not recommendable. APEX is located at an average surface pressure of 555 [mbar] while LHATPRO is at an average surface pressure of 742 [mbar], and none of these pressures are matched by the surface pressure of MERRA-2. This pressure difference is of major relevance in PWV studies due to the exponential decrease of water vapor with altitude.

Therefore, this surface pressure divergence may contribute to an overestimation and a possibly systematic offset while characterizing the PWV. Additionally, the latitudinal-longitudinal grid of MERRA-2 has a coarse resolution, preventing its

direct usage for the determination of particular locations. The closest MERRA-2 grid point for LHATPRO is located at the northwest (NW) with a distance of 26.42 [km], while the one for APEX is located at the east (E) with a distance of 26.53 [km], as seen in Figure 4.1.1. As a result, none of the MERRA-2 grid points is closer than 26 km from their respective site. To solve these issues, in this study, pressure-stratified data was chosen over integrated data, and the site PWV has been calculated using a tridimensional interpolation process.

The variables used from the MERRA-2 file are the specific humidity (QV), the pressure thickness (DELP) of each pressure level, and the mid-level pressure (PL), which is required to locate the vertical model-level layers with their corresponding atmospheric layers.

The precipitable water vapor (PWV) can be determined by the following equation (Li et al., 2020):

$$PWV = \frac{1}{g} \int_{p_s}^{p} q dp \tag{3.1.1}$$

where q is the specific humidity and g is the gravitational acceleration of the Earth. Specific humidity values (QV) along with the pressure thickness (DELP) are used to retrieve the PWV per layer; summing these values over all layers provides the total value of integrated PWV at the surface pressure levels defined by the model.

Since the native grid points of the MERRA-2 reanalysis data do not match the coordinates or pressure level of the sites under investigation on its native grid, the interpolation process is carried out, consisting of two consecutive operations:

(1) a **vertical one-dimensional interpolation** throughout the pressure layers, with the goal of determining the cumulative water vapor content at the exact pressure of the APEX or LHATPRO site, disregarding any additional component due to altitude misalignment

(2) a **geographical two-dimensional linear interpolation**, taking the pressure corrected PWV values from the closest grid points surrounding the site location and their respective distances to obtain the integrated PWV value at the site's surface pressure on its geographical location.

In this study, these interpolated values are compared to the measured PWV values obtained by the microwave radiometers of APEX and LHATPRO.

The number of closest grid points selected for APEX and LHATPRO are 6 and 4, respectively. This difference in the number of grid points for interpolation is due to the fact that APEX's latitude (-23.006°) is similar to the latitude of two of the MERRA-2 grid points (-23°). This would make the bidimensional interpolation biased towards the points on that specific latitude.

The coordinates of each grid point are shown in Figure 4.1.1 and indicated in Table 4.1.1. For future works, the Python code developed to automatically evaluate the PWV of sites using MERRA-2 data has been configured to use the 9 closest grid points to avoid a disparity on the number of points used [1].

3.2 Statistical analysis

Several plots were created using the data obtained from the bidimensional interpolation over the closest grid points shown earlier. These plots allow for a visual interpretation of the data behavior over different statistical perspectives. Relevant numerical data is shown alongside their respective plots when necessary for better understanding of the plot itself while serving as additional criteria for the comparison process.

First, in order to have a visual appreciation of the data and identify any data voids or inconsistencies, a time plot was generated, directly putting PWV data from measurements and reanalysis over each other through time, with time in the x-axis and PWV data in the y-axis, matching site measurements data with MERRA-2 timestamps and discarding any MERRA-2 timestamp for which there's no water vapor measurements. In parallel, the difference between MERRA-2 data (interpolated) and the site's measurements, defined as $\Delta\text{PWV} = \text{PWV}_{\text{MERRA-2}} - \text{PWV}_{\text{site}}$, is shown below the temporal variation plot, aiming to provide a visual representation of the data dispersion.

A more direct comparison is done through a PWV vs. PWV plot, plotting PWV vs PWV through all of the matching timestamps between the site and MERRA-2 (interpolated), with each point corresponding to one timestamp. A linear fit

[1] Available at https://doi.org/10.5281/zenodo.8221438

was applied using robust linear regressor (using SciPy's `HuberRegressor` library, which implements the Huber loss function, Huber 1964) to determine the general trend and minimize outliers effects. No offset was considered when applying the robust linear regressor due to inconsistencies between input data and the final linear fit generated when considering the presence of an offset.

Additionally, both Pearson and Spearman coefficients were calculated to quantitatively see how MERRA-2 data could linearly adjust to the site's measurements and how monotonic the correlation would be, respectively. Alongside the PWV vs. PWV plot, a table with the slope (m) value for the linear fits as well as the respective Pearson and Spearman values is shown.

A series of histograms, based on the previously defined difference ΔPWV, are generated, to show the distribution of the differences. These histograms are intended to identify any possible inconsistencies from the interpolated result. Along with them, the mean, median, and standard deviation are also provided for easier analysis and comparison.

Finally, a cumulative distribution function (CDF) plot is generated, showing the probability distribution of the PWV values overtime regardless of the timestamp. Additionally, a CDF difference plot is also provided to obtain a visual representation of the MERRA-2's CDF dispersion in comparison to the site's measurements CDF.

A summary of this process is shown in Figure 3.2.1. The specific data inputs from MERRA-2 were **specific humidity** (QV) and **pressure thickness** (DELP), needed for PWV calculation. MERRA-2 data was converted and interpolated both vertically and geographically in parallel to measurement data being time-averaged matching MERRA-2's timestamps. Once processed, MERRA-2 data was compared with measurement data in four plots:

- A **time plot**, visually showing temporal variation and data voids.
- A **PWV vs. PWV plot**, fitting a robust linear fit and obtaining the **slope** (m), the **Pearson coefficient** (r) and the **Spearman coefficient** (ρ).
- A **difference histogram**, showing the data difference dispersion and obtaining the **mean** value (μ), the **median** value and the **standard deviation** value ($sigma$) for every grid point as well as the interpolated

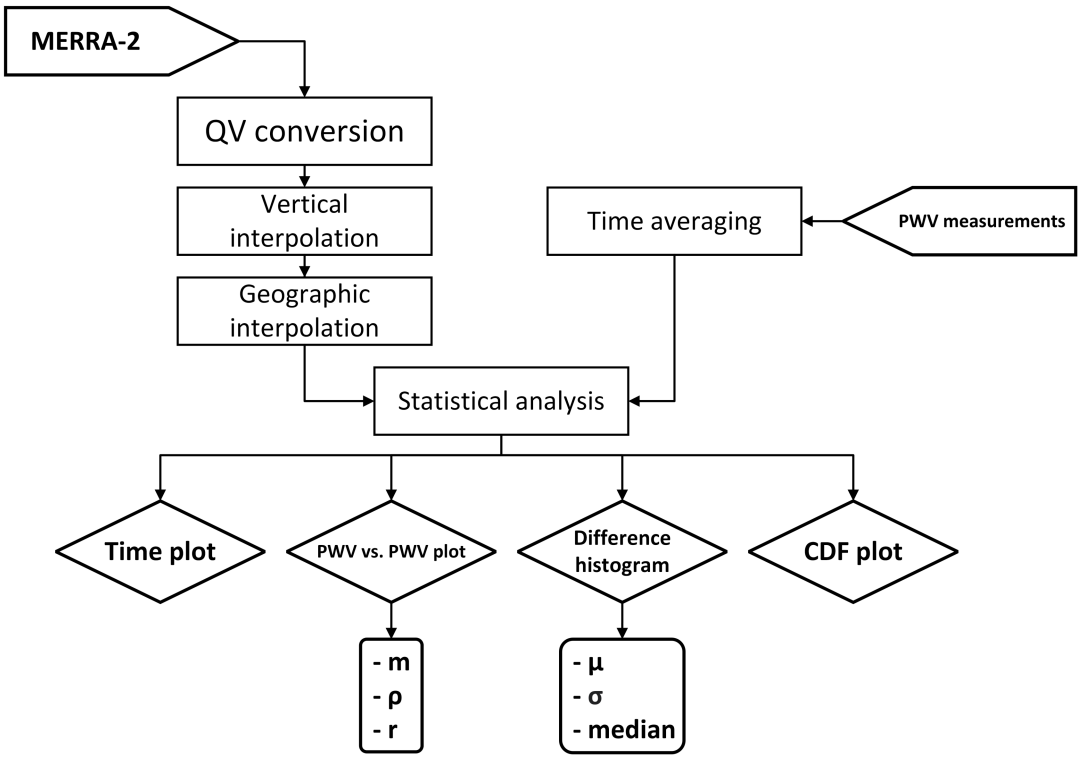


Figure 3.2.1: Illustrative block diagram summarizing the validation process made in this study. The one end pointed rectangles represents data inputs and the rectangles represent the procedural actions applied. Diamonds represent the plots generated and the rounded rectangles represent the numerical results obtained from the respective plot

data.

- A **cumulative distribution function (CDF) plot**, showing the global distribution of atmospheric water vapor concentration overtime for every grid point as well as the interpolated data.

3.3 Radiative transfer evaluation

After the validation process is completed, to complement the site-testing evaluation, a radiative transfer analysis process is made for both sites studied (APEX and LHATPRO), as well as the three additional sites of interest for the ng-EHT: two of them located in Chile and one located in the Antarctica, indicated in Table 5.1.1. This process consists in calculating the site's atmospheric transmission given the previously explained interpolated analysis of MERRA-2 data of ozone, temperature, pressure, and verified water vapor data using the radiative transfer model *am*. As in the case of APEX, the LCO site has a very small distance between site and nearest grid point, thus 6 grid points are used to determine the atmospheric parameters for its radiative analysis. For both VALLE and GARS sites, 4 grid points are used.

3.3.1 Reanalysis data processing

The MERRA-2 data used for building the configuration files necessary for the program's execution were the pressure layers, specified by their mid-layer pressure (PL), specific humidity (QV), ozone mixing ratio (O3) and air temperature (T).

When directly extracted from MERRA-2, pressure levels are expressed in Pa, QV and O3 are expressed as a mass ratio (kg kg^{-1}) and T is expressed in K. Following one of *am*'s example model files as reference (`CambridgeMA_annual_50.amc`, available in *am*'s cookbook), it's decided to work with pressure in mb and with any atmospheric column expressed as a volume mixing ratio (vmr), hence QV and O3 values must be converted to vmr before being included in the configuration file. Given that 1[mbar] = 1[hPa] = 100[Pa], the pressure conversion is $P\,[\text{mbar}] = P\,[\text{Pa}]/100$

For this study, MERRA-2 interpolated data are used to make one configuration file for the 50 percentile of data from 2019 to 2022, organized in 35 pressure levels.

Each layer is defined by its base pressure (represented as `Pbase`), its temperature (`Tbase`), and the columns of the atmospheric species to be analyzed in the model. The columns used are:

(a) `dry_air`, predefined by the *am* software, that includes the standard concentration of the different dry air components, like CH_4, CO, CO_2, N_2O and O_2 (Paine, 2022).

(b) `o3`, obtained from the MERRA-2 O3 values, that allows to manually define the ozone concentration.

(c) `h2o`, obtained from the MERRA-2 QV values, that defines the amount of water vapor present in the layer.

On the other hand, for both QV and O3, the conversion process from mass mixing ratio to volume mixing ratio is straightforward, needing the molecular mass of dry air and the molecular mass for both ozone and water vapor (Air Resources Laboratory (ARL), 2023). The case for O3 is as follows, with $m_{\text{dry}} = 28.964$ as the molecular mass of dry air and $m_{\text{o3}} = 48$ as the molecular mass of ozone.

$$\text{O3}\,(vmr) = \left(\frac{m_{\text{dry}}}{m_{\text{O3}}}\right) \cdot \text{O3}\,(mmr) \tag{3.3.1}$$

Specific humidity (q) is defined as the ratio of water vapor mass per mass unit of the total air. This differs from mass mixing ratio (w), defined as the relation between a gas mass in a mixture relative to all of the other gases in the mixture (Wallace and Hobbs, 2006). In the case of water vapor,

$$q = \frac{m_v}{m_v + m_d} \quad , \quad w = \frac{m_v}{m_d} \tag{3.3.2}$$

Where m_v is the mass of water vapor and m_d is the mass of every other component that is not water vapor, called dry air. Due to this, a preliminary conversion process from specific humidity to water vapor mixing ratio is needed

$$w(q) = \frac{q}{1-q}, \tag{3.3.3}$$

and the rest of the process is analogous to the ozone case, with $m_{\text{h2o}} = 18.015$ as

the water vapor molecular mass and $m_{\text{dry}} = 28.964$ as the dry air molecular mass.

$$\mathrm{H_2O}\,(vmr) = \left(\frac{m_{\text{dry}}}{m_{\mathrm{H_2O}}}\right) \cdot \left(\frac{QV}{1 - QV}\right) \tag{3.3.4}$$

3.3.2 Atmospheric model application

Having the atmospheric values previously converted, the configuration file for the *am* software can be built. The *am* software's cookbook is a compilation of different demonstrative configuration files used in the *am* manual (Paine, 2022), attached along the software when downloaded, from which the `CambridgeMA_annual_50.amc` demonstrative file was extracted to use its atmospheric layering as reference for the configuration files elaborated in this work.

This reference structure consists of approximately 40 pressure layers going from 0.1 mbar to 1000 mbar with a resolution that varies depending on the atmospheric section (e.g. stratosphere, mesosphere, troposphere), being finer the lower the pressure is. Based on this reference pressure levels, a vertical interpolation process was done to MERRA-2 converted data, similar to the one described for the validation process, estimating the values for the specific pressure levels in the reference configuration file, as well as adding a final pressure level corresponding to the surface layer of the respective site, given that none of the sites' surface pressures are equal to any of the levels in this reference structure. This would reduce the original 72 levels provided by MERRA-2 to roughly $30 - 35$ levels in the configuration file to be used in the *am* software. The structure of a `.amc` configuration file is composed of two main parts: a preamble and a set of layers, each one with their respective features. The preamble defines the inputs and outputs of the software when executed as shown in Figure 3.3.1. Inside the configuration file and specified with a numbered locator, the inputs are:

- (`%1`) the start frequency along with (`%2`) its unit of measure
- (`%3`) the end frequency along with (`%4`) its unit of measure
- (`%5`) the frequency step along with (`%6`) its unit of measure
- (`%7`) the zenith angle (set by default in 0°) along with (`%8`) its unit of measure

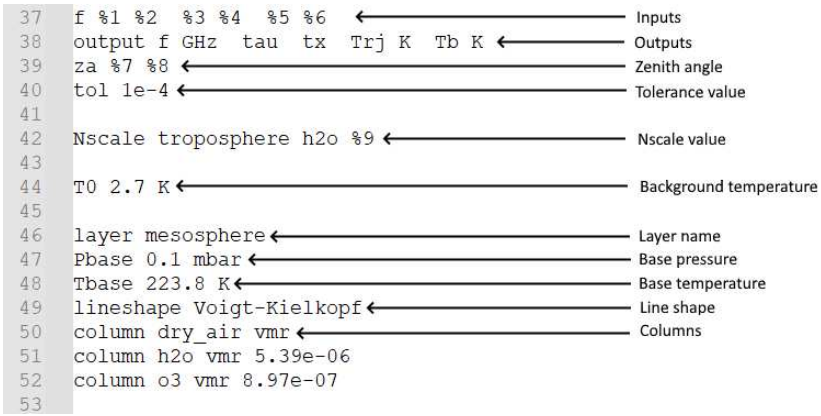


Figure 3.3.1: Extract from the `CambridgeMA_annual_50.amc` configuration file's preamble and first layer.

- (`%9`) the scale factor (dimensionless) for water vapor (set by default in 1).

The outputs shown in the example are:

- Frequency (`f`) in GHz (`GHz`)
- Opacity (`tau`) in neper (not specified, set by default)
- Transmittance (`tx`), dimensionless
- Rayleigh-Jeans temperature (`Trj`) in K (`K`)
- Brightness temperature (`Tb`) in K (`K`)

In this case, the outputs will only be transmittance and frequency in GHz. A single transmittance vs. frequency plot was generated from 200 GHz to 400 GHz with a resolution of 0.1 GHz using the data obtained with the *am* software for the 5 sites previously mentioned, each with its respective keyword and highlighting the frequencies of interest for the ng-EHT: 230 and 345 [GHz].

A summary of this process can be seen in Figure 3.3.2. THe specific data inputs from MERRA-2 were **specific humidity** (QV), **ozone mass mixing ratio** (O3) and **air temperature** (T). The input data was interpolated vertically to the reference pressure labels and geographically to the respective site location. The 50th percentile was extracted for every value and in QV and O3's case, a conversion process was required to obtain both water vapor volume mixing ratio and ozone volume mixing ratio, respectively. A transmittance vs. frequency plot was generated using the output data obtained with the *am* software for each site of interest.

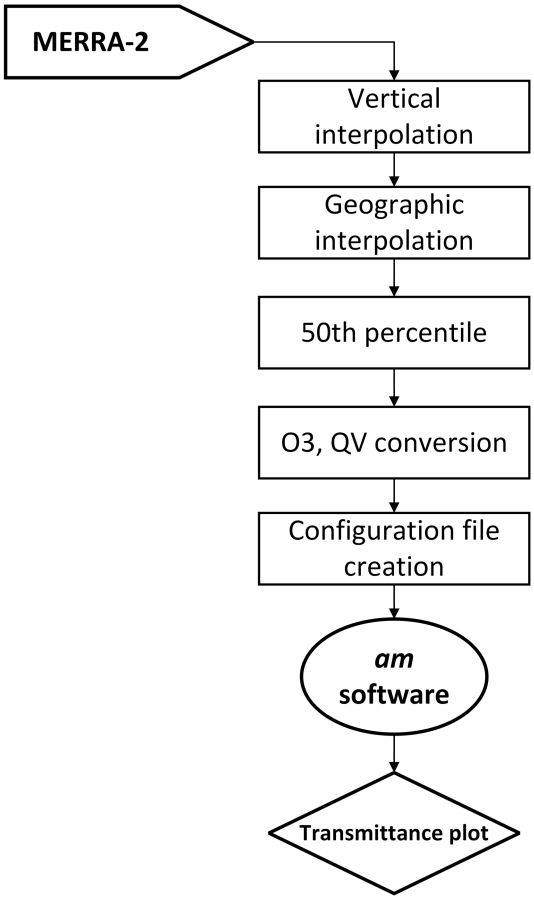


Figure 3.3.2: Illustrative block diagram summarizing the atmospheric performance evaluation. The one end pointed rectangle represents data input and the rectangles represent the actions applied. The oval shape represents the software implemented and the diamond shape represents the plot output.

Chapter 4

Results and discussion

4.1 Data sources

For the validation process, atmospheric data obtained from the Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2) reanalysis database is compared with water vapor measurements extracted from two astronomical sites located in Chile: Atacama Pathfinder Experiment (APEX) in Chajnantor, Atacama, and Cerro Paranal's Low Humidity and Temperature Profiling (LHATPRO) microwave radiometer, near Antofagasta, Chile.

4.1.1 Water vapor measurements

The Atacama Pathfinder EXperiment (APEX) (Güsten et al., 2006) is a 12-m submillimeter telescope located at 5107 m altitude on Llano de Chajnantor in the Chilean High Andes (-23.006°, -67.759 °), which works as the pathfinder for other sub-mm missions, specifically the Atacama Large Millimeter Array (ALMA), from which the telescope is a modified prototype made for single dish operations. A 183 GHz water vapor radiometer built by Radiometer Physics GmbH (MeckenHeim, Germany) is located in the Cassegrain cabin of the APEX telescope and provides PWV observations since 2006. The measurements are publicly available through the APEX Weather Query Form[1]. In this data analysis process, an averaging process was applied to the raw data to match with MERRA-2 timestamps (obtained every 3 hours) by averaging all the data with a temporal

[1] http://archive.eso.org/wdb/wdb/eso/meteo_apex/form

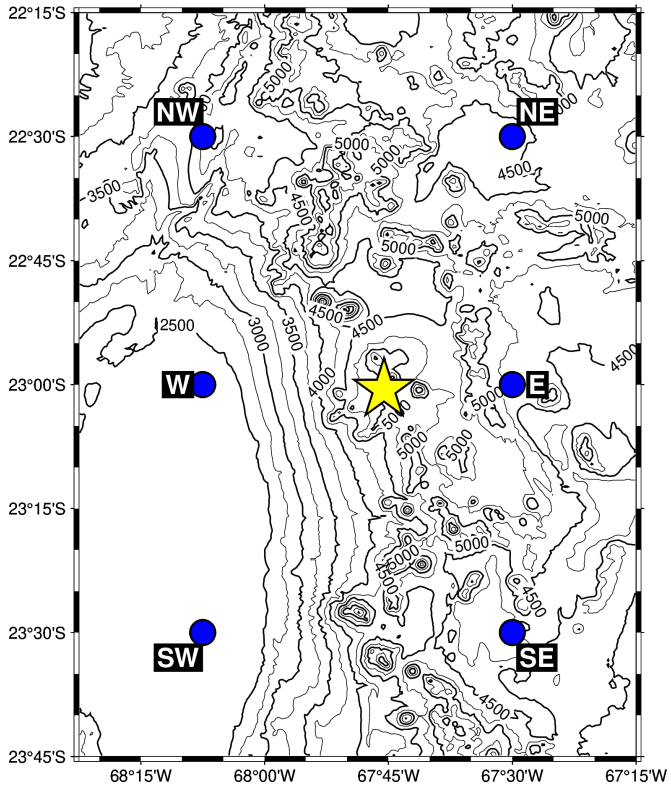
	APEX		LHATPRO	
	Latitude (deg)	Longitude (deg)	Latitude (deg)	Longitude (deg)
Instrument Site Location	−23.008	−67.759	−24.627	−70.404
MERRA-2 Northwest (NW)	−22.5	−68.125	−24.5(∗)	−70.625(∗)
MERRA-2 Northeast (NE)	−22.5	−67.5	−24.5	−70
MERRA-2 Southwest (SW)	−23.5	−68.125	−25	−70.625
MERRA-2 Southeast (SE)	−23.5	−67.5	−25	−70
MERRA-2 West (W)	−23	−68.125		
MERRA-2 East (E)	−23(∗)	−67.5(∗)		

Table 4.1.1: Location of the PWV radiometer instruments of APEX (Cerro de Chajnantor, Chile) and LHATPRO (Cerro Paranal, Chile) and the surrounding MERRA-2 grid points evaluated. Locations indicated with (*) correspond to the nearest grid points.

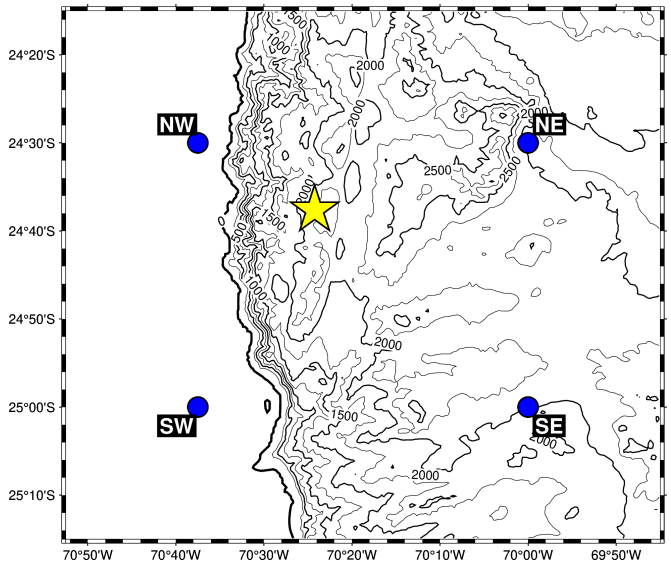
window of ± 1.5 hrs from it..

The Low Humidity and Temperature Profiling (LHATPRO) (Kerber and Querel, 2014) is a microwave radiometer used to monitor sky conditions over ESO's Paranal observatory (-24.627°, -70.404°), in Cerro Paranal, Antofagasta, Chile at an altitude of 2635 meters. It consists of a humidity profiler (183 GHz) that measures PWV in the ranges of 0-20 mm which can also obtain temperature profiles (51-58 GHz) and an infrared camera (∼ 10 micrometers) for cloud coverage detection. This instrument was commissioned between October and November 2011. Its measurements started from January 2014 and are publicly available through the LHATPRO Query Form[2]. The same averaging process applied to APEX was applied to the raw data provided by the LHATPRO query, to match the LHATPRO timestamps with MERRA-2 timestamps.

In this study, the data analyzed at both sites corresponds to a period of 4 years, which ranges from 2019-01-01 to 2022-12-31. It is relevant to highlight that during this period only weak-moderate el Niño climatic events have been observed. Additionally, a set of incorrectly calibrated PWV data from APEX was not included in the analysis process and negative PWV values were filtered from LHATPRO.

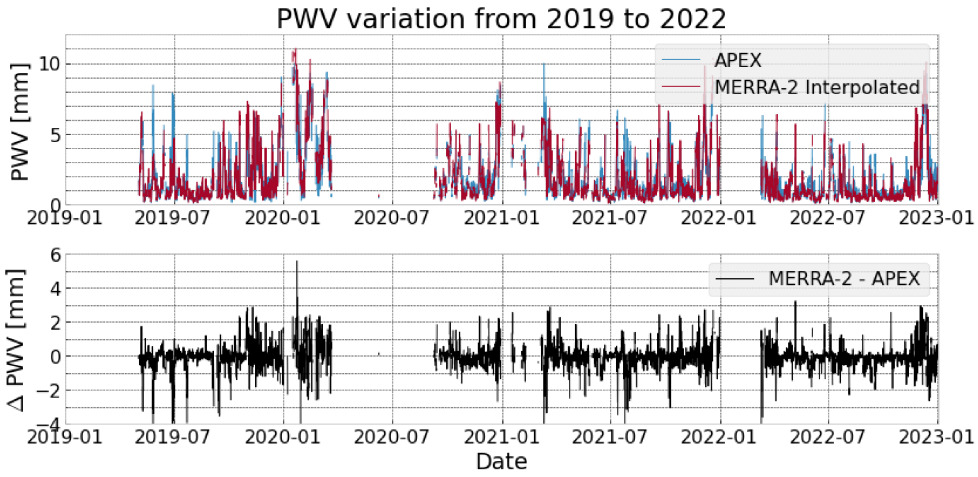


(a) APEX (Cerro Chajnantor, Chile)

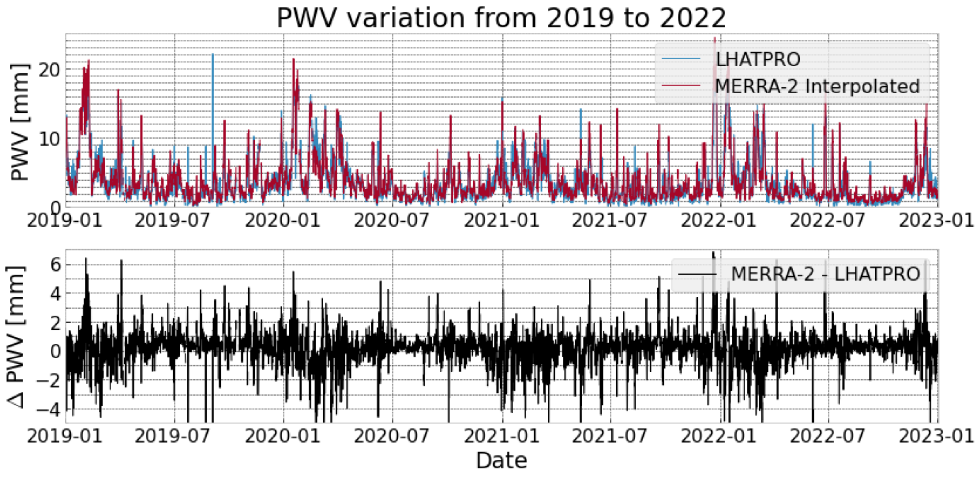


(b) LHATPRO (Cerro Paranal, Chile)

Figure 4.1.1: Topographic maps of the locations indicated in Table 4.1.1. The yellow star represents the location of a) APEX and b) LHATPRO sites and the blue circles indicate the MERRA-2 grid points, labeled according to their position with respect to the site



(a) APEX (Llano de Chajnantor, Chile)



(b) LHATPRO (Cerro Paranal, Chile)

Figure 4.2.1: Temporal plots of PWV for (a) APEX and (b) LHATPRO study sites through the years 2019 to 2022. The upper plot shows (in blue) the radiometric measurements and (in red) the MERRA-2 bidimensionally interpolated data. The lower plots (black line) show the numerical differences between the MERRA-2 interpolated estimate and the site measurements.

	m	r	ρ
APEX	0.946	0.915	0.901
LHATPRO	0.927	0.928	0.906

Table 4.2.1: Values of the slope (m) of the linear fit (using a robust linear regression) and the Pearson (r) and Spearman (ρ) correlation coefficients obtained by the comparison to the MERRA-2 interpolated data at both study sites.

4.2 Statistical analysis

4.2.1 Time plot

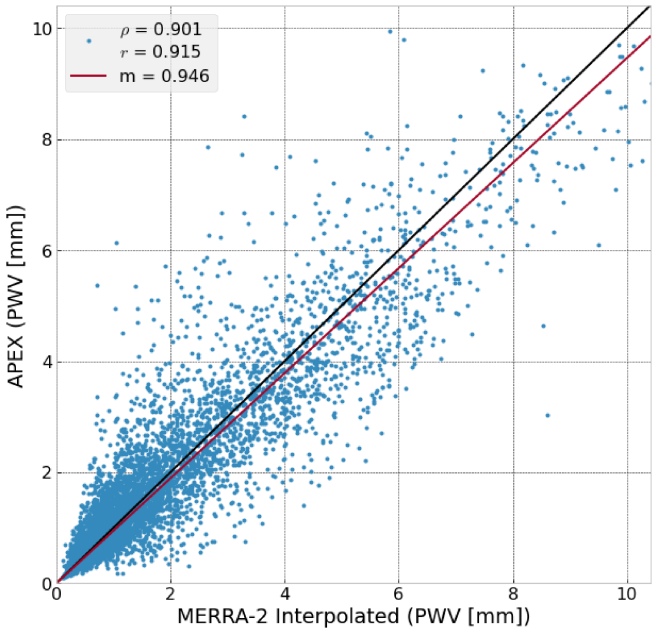
Figure 4.2.1 shows the time plot obtained during the study period for both sites. At APEX, the time interval starts in June of 2019, and it has visible data voids during half of 2020, the first quarter of 2021, and the first quarter of 2022. These data gaps are attributed to the operation of the radiometer, which measures only when observations are made and the telescope's shutter is open. It is relevant to note that this season period corresponds to the so-called Andes altiplanic winter (Otarola et al., 2019) which generates a large increase in PWV levels and prevents astronomical observations. On the other hand, LHATPRO (shown in Figure 4.2.1b) does not present large data gaps because its operation is independent on astronomic observations.

The time plots shown in this figure serve as a visual representation of the data dispersion along the years. Additionally, these plots allow an easier appreciation of the measurement data outliers, mainly attributed to shutdown of the device due to extreme weather conditions and/or technical issues. It is relevant to highlight that these plots show that the MERRA-2 interpolated data provides a similar PWV response to the radiometric measurements, as observed in the PWV difference plots.

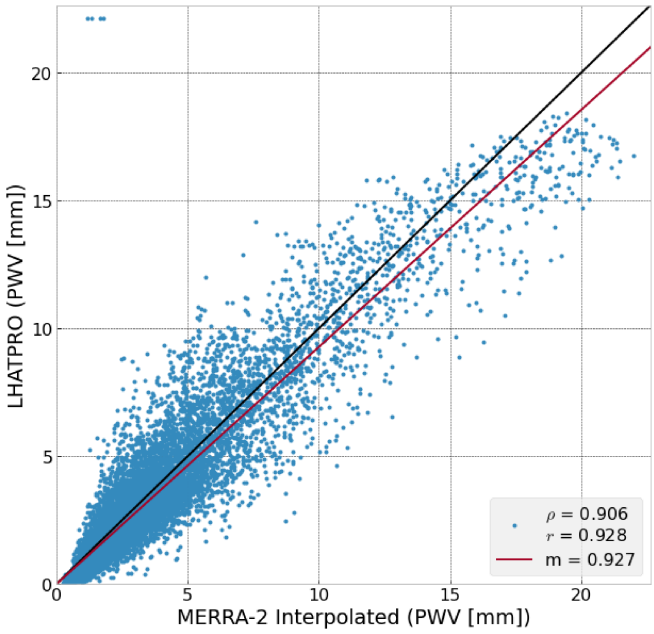
4.2.2 PWV vs. PWV plot

The results suggest that both sites show a quite good correlation that is linear and monotonous based on the calculated correlation coefficients. The PWV vs. PWV plots shown in Figure 4.2.1 represent the relationship between the PWV data at equal timestamps for MERRA-2 and measurements at each respective site.

[2] http://archive.eso.org/wdb/wdb/asm/lhatpro_paranal/form



(a) APEX (Llano de Chajnantor, Chile)



(b) LHATPRO (Cerro Paranal, Chile)

Figure 4.2.2: PWV vs. PWV plots (blue dots) comparing the site measurements and the data obtained by the interpolation of MERRA-2 data at a) APEX and b) LHATPRO sites. The red line represents the robust linear regression and the black line represents the 1:1 line. Legend values indicate the Spearman correlation coefficient (ρ), the Pearson correlation coefficient (r), and the slope of the linear regression (m, defined by the equation $\mathrm{PWV}_{\mathrm{site}} = m \cdot \mathrm{PWV}_{\mathrm{MERRA\text{-}2}}$).

Additionally, a robust linear fit and the 1:1 line are plotted. The overall response of the MERRA-2 interpolated data at both sites and their corresponding linear fits showed a slope greater than 0.9, ($m_{\text{APEX}} = 0.946$ and $m_{\text{LHATPRO}} = 0.927$). Additionally, the Pearson correlation coefficient (r) is greater than 0.9 in both cases ($r_{\text{APEX}} = 0.915$ and $r_{\text{LHATPRO}} = 0.928$). Lastly, the Spearman coefficient, aimed to measure monotonic correlations regardless of their linearity, resulted in $\rho_{\text{APEX}} = 0.901$ for APEX's case, and $\rho_{\text{LHATPRO}} = 0.906$ for LHATPRO's case. For convenience, these six values are shown in Table 4.2.1.

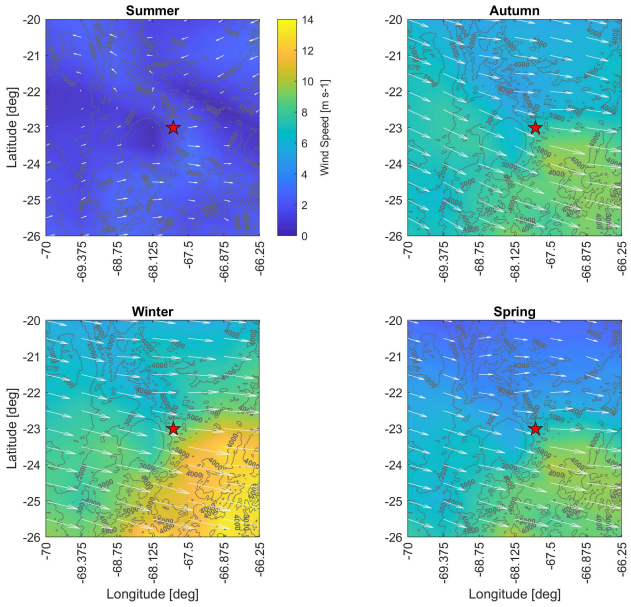
4.2.3 Difference histogram plot

The histogram of the difference between the MERRA-2 interpolated data and the site's measurements is provided in Figure 4.2.5 for each of the study sites along with histograms of each of the MERRA-2 grid points used for the interpolation. The mean, median, and standard deviation of these differences are listed in Table 4.2.2.

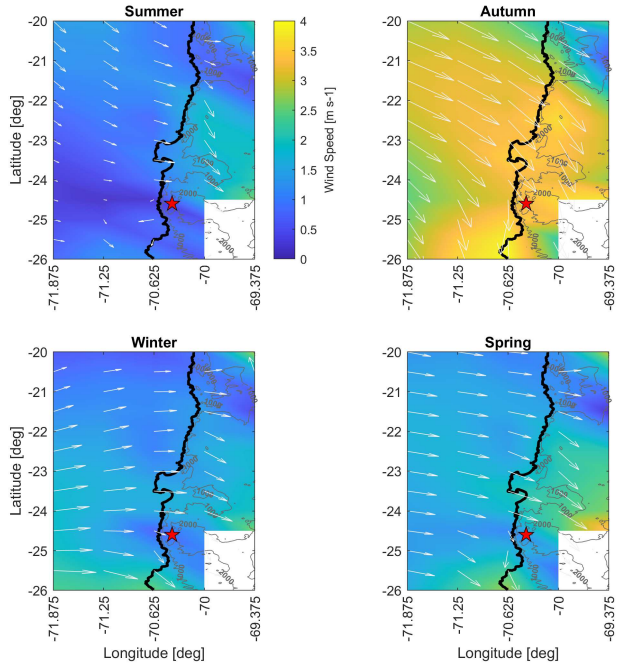
For APEX, the difference histograms for all grid points, and the interpolated values, are highly similar as seen in Figure 4.2.5a. Alternatively, in the LHATPRO's case, shown in Figure 4.2.5b, the histograms obtained at NE and SE points have their central value clearly offset towards positive values. On the other hand, the difference histograms for the NW and SW grid points, and the interpolated data, while more centered, are still not centered on zero.

Looking at the statistical values shown in Table 4.2.2, it can be seen that, in the APEX study case, the interpolated data has the lower value for σ and the μ value closest to 0. Nevertheless, in this case, the median value closest to zero is obtained by the data from the SE grid point with $\text{median}_{\text{SE}} = -0.004$, followed by the interpolated point with $\text{median}_{\text{MERRA-2}} = -0.007$. On the other hand, in the LHATPRO's dataset, SE has the lowest value for σ, NW has the lowest value for μ, and SW has the lowest median value.

The results obtained at APEX suggest a very good agreement between the MERRA-2 reanalysis model interpolation method and the measurements. Nevertheless, at LHATPRO, larger differences are obtained. These discrepancies of the method can be explained by the relatively large histogram center offset at 2 of the 4 grid points used (NE and SE), which has biased the interpolation process. The cause

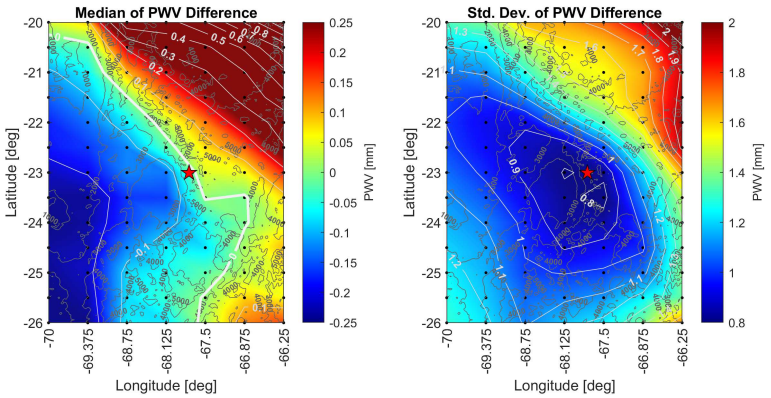


(a) APEX (Llano de Chajnantor, Chile)

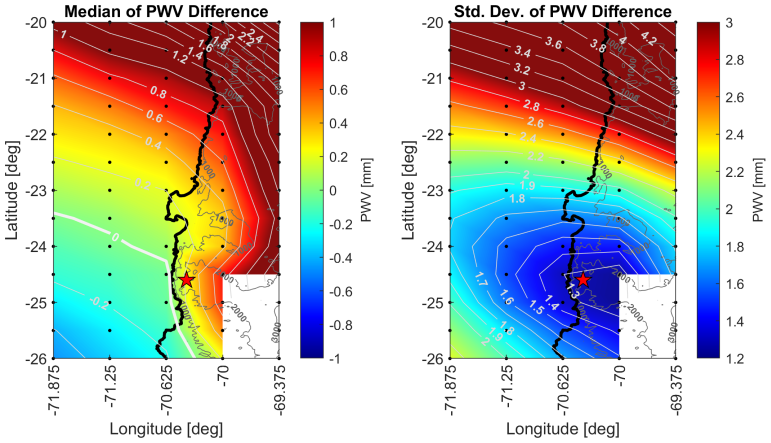


(b) LHATPRO (Cerro Paranal, Chile)

Figure 4.2.3: MERRA-2 winds (a) at the pressure of APEX site (555 mbar) and (b) at the pressure of LHATPRO site (742 mbar) at different seasons of 2019. The color gradient represents the wind speed magnitude, while the arrows represent the direction of the wind. The red star indicates the respective site location.



(a) APEX (Llano de Chajnantor, Chile)



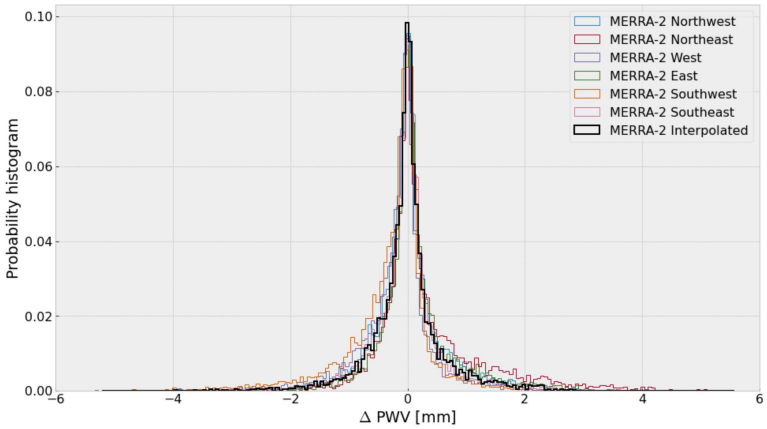
(b) LHATPRO (Cerro Paranal, Chile)

Figure 4.2.4: Maps of median and standard deviation of the difference between site measurements and MERRA-2 values (a) at the pressure of APEX site (555 mbar) and (b) at the pressure of LHATPRO site (742 mbar) during the year 2019. The black dots represent MERRA-2 grid point locations and the red star indicates the respective site location.

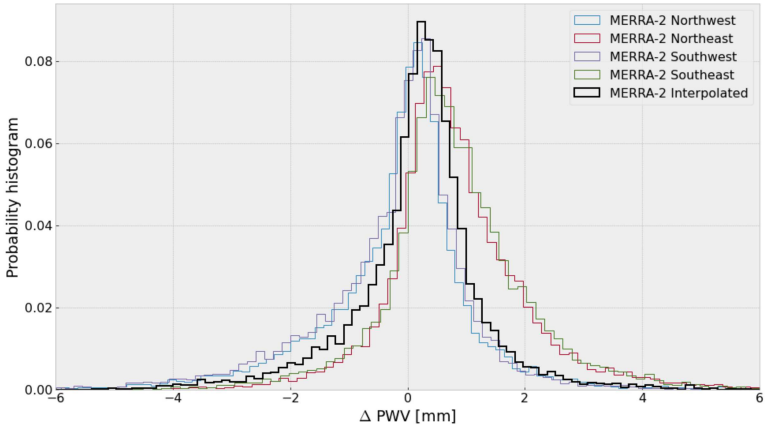
of these offsets is yet unknown, but an initial analysis suggests that it can be attributed to the large variability of the winds present in that area, close to the Pacific Ocean. Figure 4.2.3 shows the average wind velocity and direction during different seasons at the same pressure level of each of the study sites obtained from the MERRA-2 reanalysis model dataset for the year 2019. In particular, Figure 4.2.3a shows the winds present at the pressure of the APEX site. Very strong winds, up to 14 meters/sec, with a regular direction pattern towards the south-east are present during all seasons except summer. In the summer period, which is the period of higher water vapor concentration, the wind direction is not well defined but the average wind speed is significantly lower.

On the other hand, Figure 4.2.3b shows the winds present at Cerro Paranal, Antofagasta, Chile. In this case, the strongest winds are found during autumn but these have lower speeds than in the case of APEX (i.e., 4 meters/sec maximum wind speed). In general, the overall winds at the LHATPRO pressure level head towards east, but large directional changes can be seen near the location of the site. This effect can be associated to what was shown in the histograms of Figure 4.2.5b, in which only the eastward points (NE and SE) present their distribution offset towards positive values. This effect can be attributed to wind interactions with the topography of the area. It is suggested that LHATPRO shows a response similar to western grid points because it is located on the summit of Cerro Paranal, where eastern winds have no interaction with the surrounding topography. Alternatively, the eastern grid points are impacted by the wind changes produced by the nearby mountain structures. Therefore, the coarse spatial resolution of the MERRA-2 grid does not provide enough representability of the distribution of the atmospheric parameters found within this area. Note that, in the APEX case, surrounding mountains have lower altitude than the observatory, but much higher topographic elevations are obtained at the LHATPRO pressure.

Figure 4.2.4 shows the geographical distribution for both median and standard deviation (σ) of the difference between the data obtained from MERRA-2 and the site measurements. Figure 4.2.4a shows that the APEX site is near the median zero value (represented by the white line) and also near the area with the lowest standard deviation values. Figure 4.2.4b shows the LHATPRO case. In this case, the median zero value white line is closer to SW and NW, located west from the site. On the other hand, the standard deviation of the distribution is similar in



(a) APEX (Llano de Chajnantor, Chile)



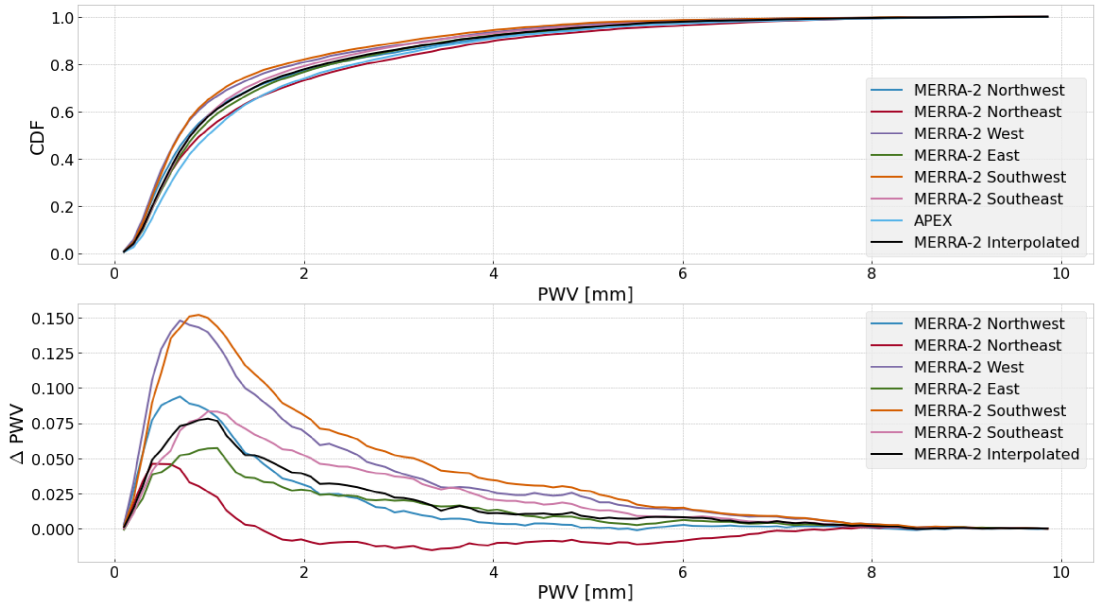
(b) LHATPRO (Cerro Paranal, Chile)

Figure 4.2.5: Difference histograms for every MERRA-2 grid point for both (a) APEX and (b) LHATPRO sites, along with the interpolated data over both sites. The difference is defined as $PWV_{\text{MERRA-2}} - PWV_{\text{site}}$.

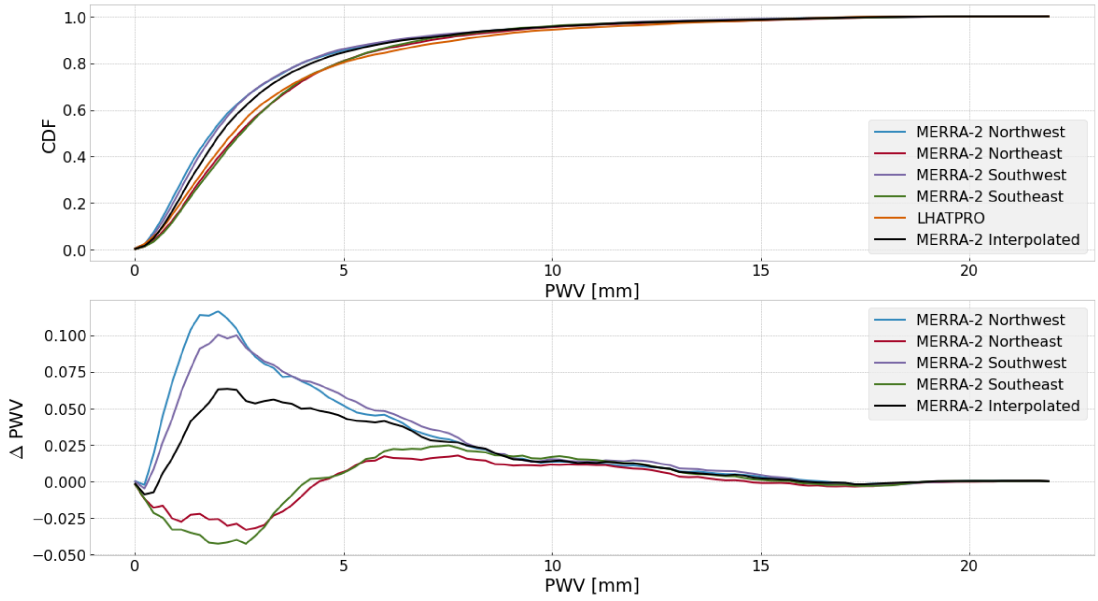
the NE, NW, and SE grid points, indicating that they have similar variabilities. These results also support the assumption that the distribution is biased due to the topographic impact on the winds within the area.

4.2.4 Cumulative distribution function (CDF) plot

To provide a useful measure of the site observability, the cumulative distribution function (CDF) is analyzed in both study sites. Figure 4.2.6 shows the data distribution without constraining it to time-stamp matching, ensuring the usage of most of the radiometric measurements. Difference CDF plots are also provided to help identify the differences and similarities obtained by the distribution at the two study sites.



(a) APEX (Llano de Chajnantor, Chile)



(b) LHATPRO (Cerro Paranal, Chile)

Figure 4.2.6: CDFs obtained from 2019 to 2022 by the radiometric measurements at a) APEX and b) LHATPRO, the nearby MERRA-2 grid points indicated in Table 4.1.1, and the MERRA-2 bidimensionally interpolated data (black line). The bottom plot shows the difference between the site measurements and the corresponding MERRA-2 model CDF values

	APEX			LHATPRO		
	σ [mm]	Median [mm]	μ [mm]	σ [mm]	Median [mm]	μ [mm]
Bidimensional Interpolation	0.68	-0.007	-0.02	1.240	0.256	0.171
MERRA-2 Northwest (NW)	0.76	-0.021	0.007	1.399	0.021	-0.169
MERRA-2 Northeast (NE)	0.969	0.062	0.311	1.352	0.68	0.827
MERRA-2 Southwest (SW)	0.722	-0.125	-0.309	1.412	0.006	-0.234
MERRA-2 Southeast (SE)	0.703	-0.004	-0.041	1.406	0.703	0.807
MERRA-2 East (E)	0.753	0.033	0.117			
MERRA-2 West (W)	0.720	-0.094	-0.207			

Table 4.2.2: Statistical values (i.e., standard deviation, σ, median, and mean value, μ) of the difference between measured radiometric PWV values and every MERRA-2 grid point indicated in Table 4.1.1, along with the interpolated data, for both study sites during 2019.

In APEX's case, shown in Figure 4.2.6a, all MERRA-2 grid points and the interpolated data, present some degree of overestimation compared to the site measurements. This effect can be seen in the difference plot, where all of the curves show positive values. It is important to note that NE and E both provide the most similar distribution to measurements, followed by the interpolated data.

On the other hand, in LHATPRO's case, both NW and SW present an overestimation of PWV in comparison to the site's data, noticeable in both the CDF plot and the CDF difference plot in Figure 4.2.6b. Meanwhile, both NE and SE show a similar distribution compared to the site's data, as appreciated in the CDF difference plot. However, the interpolated data CDF has a rather noticeable deviation from the site data, specifically between 0 and ~ 3.5, most likely due to the strong deviation that 2 of the 4 grid points have previously stated. This strong offset would unavoidably apply a bias into the interpolation process and it's clearly more noticeable in the CDF difference plot. Although this result is related to the histogram difference result previously indicated, a more detailed analysis would be required to identify the origin of the MERRA-2 data disparities, which is beyond the scope of this study.

Chapter 5

Performance analysis on other sites

5.1 Sites of interest

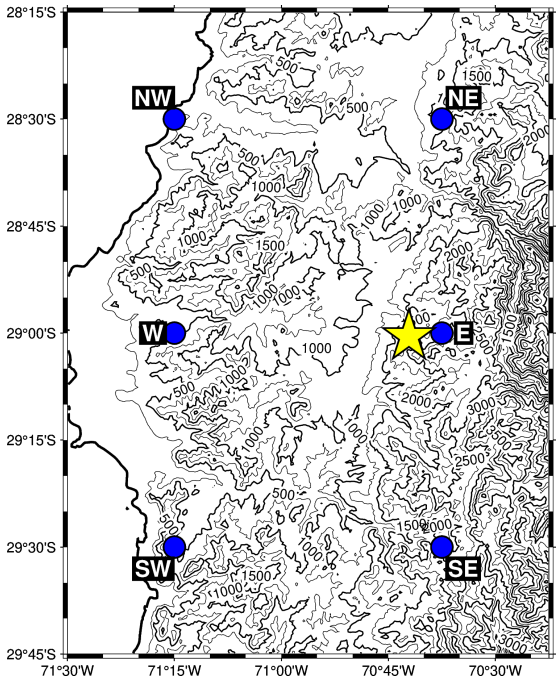
Raymond et al. 2021 made a brief atmospheric evaluation over 45 potential new sites for the ng-EHT across the world. These potential sites were chosen based whether they're located in high non-equatorial peaks or already known low-PWV areas (e.g. Alps, Andes, The Himalayas); or if they provide good Fourier coverage with current EHT sites. Among these sites, three were chosen for this study: two located in Chile and one located in Antarctica. For each of these sites, as well as for both sites previously studied for the validation process, an atmospheric performance evaluation is done, obtaining the global atmospheric transmittance profile over the span of four years through radiative transfer models.

Las Campanas Observatory (LCO) is an observatory located in the Andes of the Atacama region (-29.008°, -70.702°). Its location and six closest MERRA-2 grid points are shown in Figure 5.1.1.

General Bernardo O'Higgins base (GARS) is located in Antarctica (-63.321°,

	Latitude (deg)	Longitude (deg)	Elevation (m)
Las Campanas Observatory (LCO)	−29.008	−70.702	∼ 2300
Valle Nevado (VALLE)	−33.353	−70.249	∼ 3000
General Bernardo O'Higgins Base (GARS)	−63.321	−57.9	∼ 100

Table 5.1.1: Geographic location of the three additional sites of interest for the ng-EHT: two in Chile and the third one in Antarctica. These sites will be analyzed along with APEX and LHATPRO for their transmittance profiles



(a) Las Campanas Observatory (Atacama, Chile)

Figure 5.1.1: Topographic map of the location of LCO. The yellow star represents the location of LCO and the blue circles indicate the MERRA-2 grid points, labeled according to their position in respect to the site

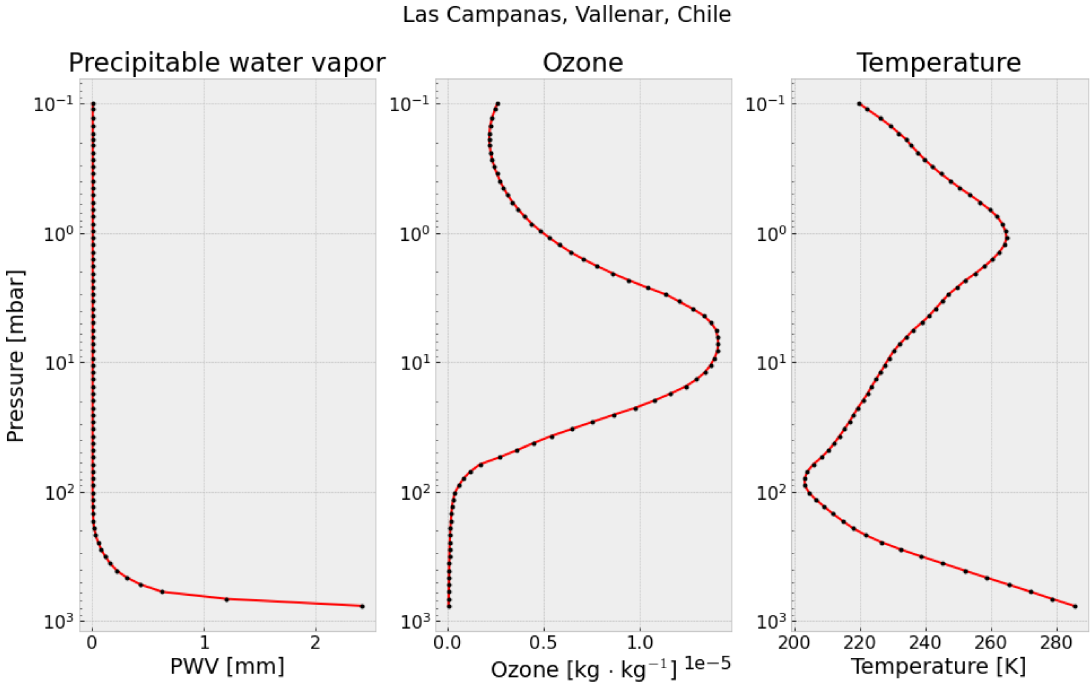
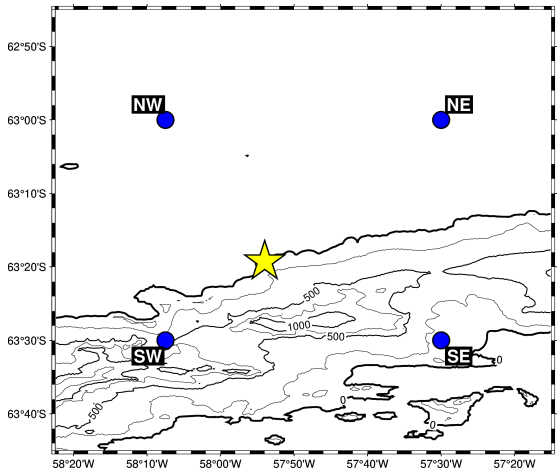


Figure 5.1.2: Mean vertical profiles over LCO from 2019 to 2022.



(a) General Bernardo O'Higgins Base (Antarctica)

Figure 5.1.3: Topographic map of the locations of GARS. The yellow star represents the location of GARS and the blue circles indicate the MERRA-2 grid points, labeled according to their position in respect to the site

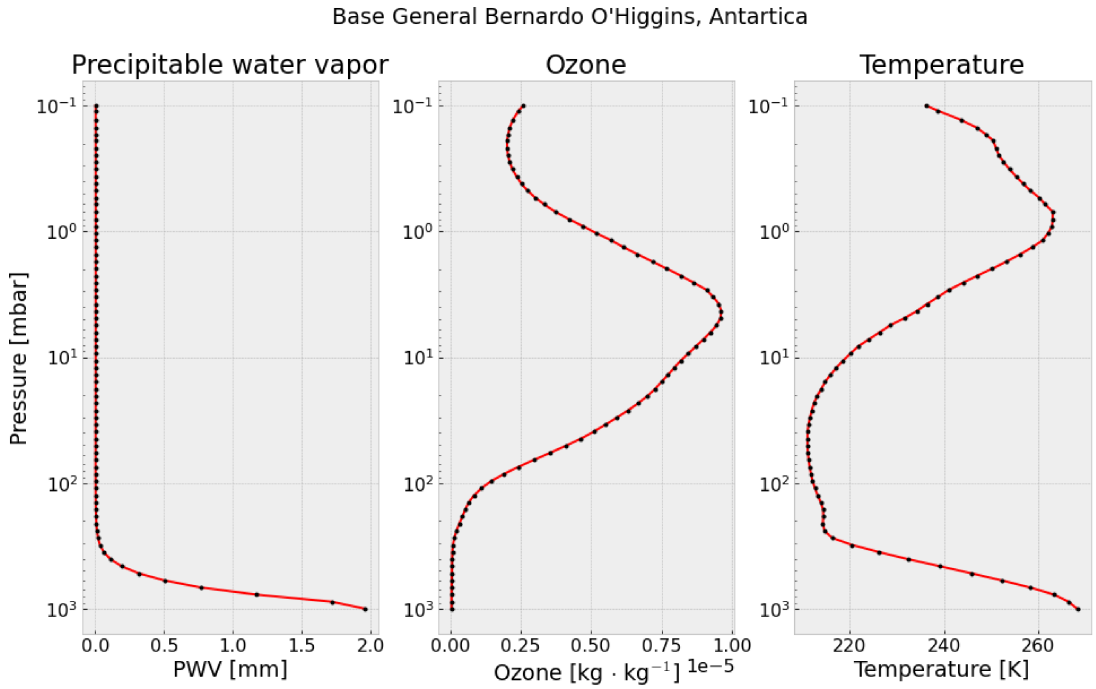
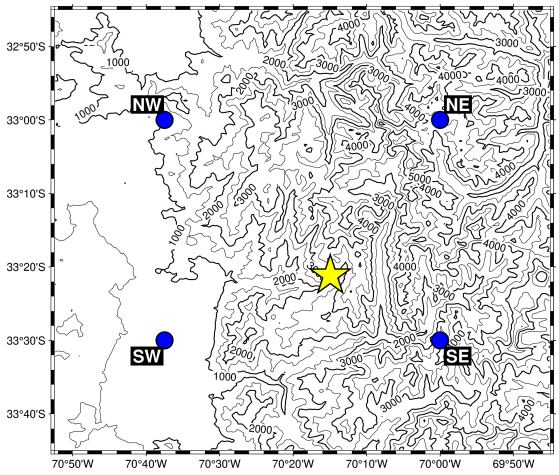


Figure 5.1.4: Mean vertical profiles over GARS from 2019 to 2022.



(a) Valle Nevado (Santiago, Chile)

Figure 5.1.5: Topographic map of the location of VALLE. The yellow star represents the location of VALLE and the blue circles indicate the MERRA-2 grid points, labeled according to their position in respect to the site

	APEX	LHATPRO	VALLE	LCO	GARS
Transmittance at 230 [GHz]	0.936	0.872	0.841	0.81	0.516
Transmittance at 345 [GHz]	0.793	0.614	0.539	0.464	0.096

Table 5.2.1: Transmittance values at 230 and 345 [GHz] for all 5 of the sites of interest.

-57.9°). Its location and four closest MERRA-2 grid points are shown in Figure 5.1.3.

Valle Nevado is a ski resort located 46 km east from Santiago de Chile (-33.353°, -70.249°) with an average altitude of ∼3000 km from lowest to highest point. Its location and four closest MERRA-2 grid points are shown in Figure 5.1.5.

Pressure vertical profiles using the mean values of PWV, ozone and temperature are generated using MERRA-2 data over the span of 4 years, shown in Figs. 5.1.2, 5.1.4 and 5.1.6.

5.2 Transmittance forecast

The atmospheric transmittance is evaluated using the *am* model over the two study sites and also over the three sites in Chile with interest for the ng-EHT project mentioned in Table 5.1.1. The transmittance is evaluated from 200 GHz to 400 GHz, covering the two frequencies of interest for the ng-EHT: 230 GHz

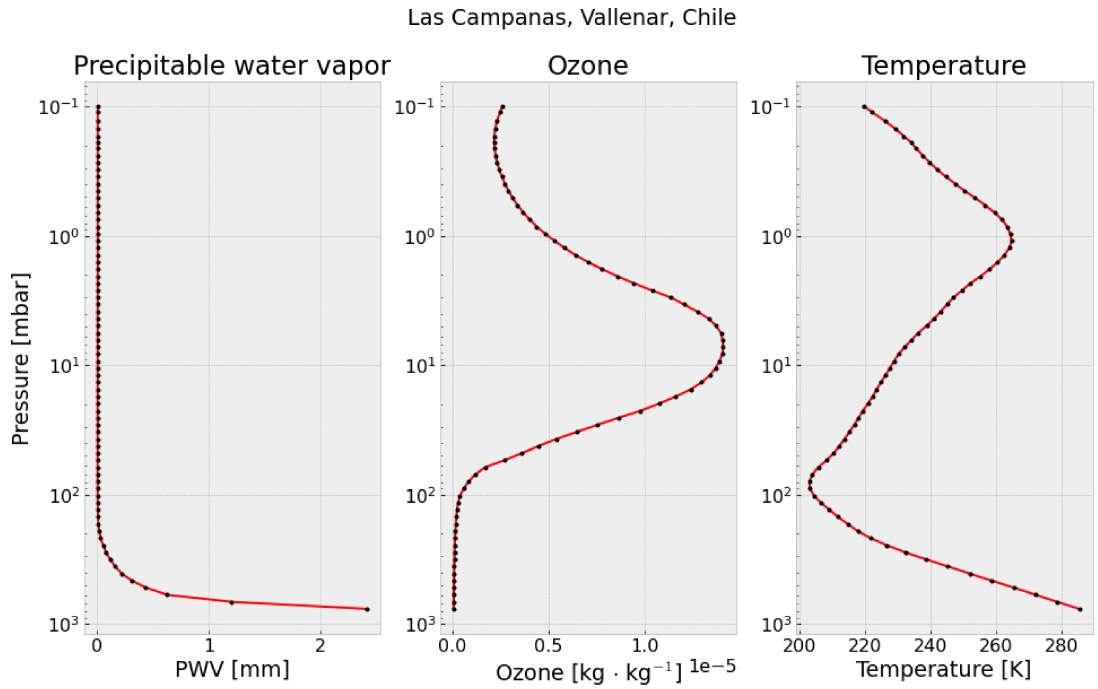


Figure 5.1.6: Mean vertical profiles over VALLE from 2019 to 2022.

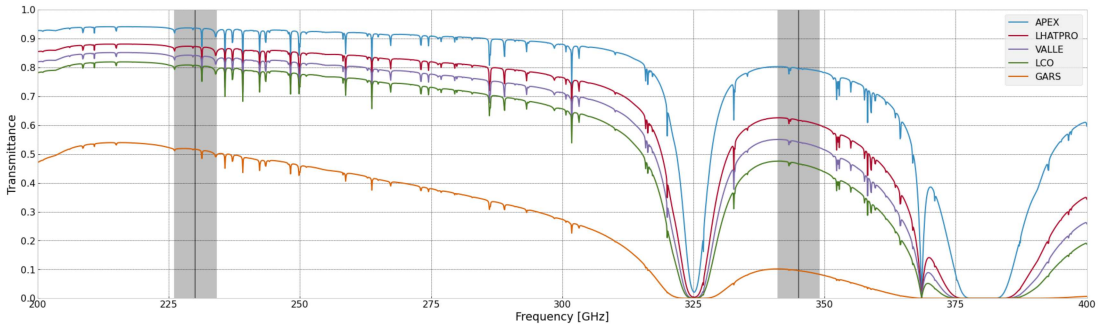


Figure 5.2.1: Transmittance profiles for the 50 percentile, obtained with the *am* software for both APEX and LHATPRO, as well as the three new sites of interest for the ng-EHT, showed in Table 5.1.1, located in Chile and the Antarctica (Raymond et al., 2021). Both black lines along with their grey bands are the two new frequencies of interest: 230 GHz and 345 GHz respectively. The bands have an arbitrary width of ±4 GHz from the frequency value of interest.

and 345 GHz. To provide a statistically significant representation of the sites, this evaluation was made based on percentile 50 of all the time-stamps available from 2019 to 2022, and the resulting plot can be seen in Figure 5.2.1. Additionally, to provide an easier visualization, the obtained values of transmittance at both frequencies of interest are also indicated in Table 5.2.1

As seen in Figure 5.2.1, APEX presents high transmittances along the considered frequency range despite the presence of the water vapor emission line at 325 GHz, the carbon monoxide (CO) emission line at 386.5 GHz, and the oxygen (O) emission line at 370 GHz. LHATPRO, VALLE, and LCO have a similar but progressively worse transmittances, most probably because of their elevation above the sea level. These three sites show a slightly steeper decrease in transmittance between 200 and 325 GHz. On the other hand, GARS (in Antarctica) shows a much lower overall performance due to its low altitude, thus presenting a much higher atmospheric water vapor content. In this case, it is not even possible to identify the CO and O lines because of the large attenuation and it has no transmittance from approximately 370 GHz onwards.

APEX, LHATPRO, and LCO sites provide results that were expected and previously studied in the literature, as observatories are already present at these locations. However, the results obtained at VALLE are promising due to their better overall transmittance when compared to LCO. Indeed, the results at VALLE are similar to LHATPRO, although their altitudes are different. This makes VALLE a promising candidate for a new ng-EHT telescope.

Chapter 6

Conclusion

In this study, a new methodology to perform site-testing using data from the MERRA-2 reanalysis model is provided. PWV data of the sites of study is obtained from the MERRA-2 dataset and interpolated in pressure and spatially to match the surface pressure and location of the site. Results from this analysis are compared to microwave radiometer water vapor measurements at two world-class astronomical sites in Chile: APEX, in Llano de Chajnantor, and LHATPRO, in Cerro Paranal. Additionally, the atmospheric mm/sub-mm transmittance is calculated using the *am* model and MERRA-2 interpolated data at five different sites in Chile with interest for the ng-EHT project.

The results obtained show that the estimation of PWV values at the study sites of APEX and LHATPRO is robust enough to complement the evaluation of the site's atmospheric quality in terms of PWV concentration over time. This method allows the analysis of an extensive amount of sites without the need for an on-site measuring system. Additionally, this method allows expanding to a full opacity evaluation system using MERRA-2, not only based on PWV but on more variables provided by the MERRA-2 dataset, such as the liquid water path or cloud coverage.

The evaluation of this new site-testing method on the proposed ng-EHT sites in the Chile showed favorable atmospheric performance in Valle Nevado, a touristic place located near Santiago de Chile, that currently does not host any astronomical observatories. This location is not only important from a scientific perspective, but from a logistic perspective as well, being Valle Nevado near the country's

capital city, that would provide easier accessibility and more manpower. This suggests the existence of more potential sites for the ng-EHT that currently do not host astronomical observations, and encourage more site-testing analyses at different locations that could be useful, disregarding the presence of observatories, to take profit of the free and open-access quality of the MERRA-2 reanalysis model dataset.

As future work, more studies can be done to provide a more robust usage of MERRA-2 as an atmospheric quality evaluation tool, reinforcing the statistical methods used and obtaining water vapor measurements from other sites. Also taking into account the already seen disparities in Paranal's case, looking into a possible cause and its respective solution/correction. In parallel, the code used for this study will be further polished to standardize the atmospheric characterization, utilizing 9 grid points instead of 4/6 to ensure a general coverage for the site of interest as well as a wider scope to look for any possible inconsistencies for the results obtained.

Bibliography

Air Resources Laboratory (ARL) (2023). Converting to Mixing Ratio ©. https://www.ready.noaa.gov/documents/Tutorial/html/chem_ratio.html. Last visited: 09-13-2022.

Bosilovich, M. G., Lucchesi, R., and Suarez, M. J. (2016). MERRA-2: File Specification.

Coakley, J. A. and Yang, P. (2014). *Atmospheric radiation: a primer with illustrative solutions*. Wiley series in atmospheric physics and remote sensing. Wiley-VCH, Weinheim.

Cortés, F., Reeves, R., and Bustos, R. (2016). Analysis of the distribution of precipitable water vapor in the Chajnantor area: PWV IN THE CHAJNANTOR AREA. *Radio Science*, 51(7):1166–1175.

Elachi, C. and Van Zyl, J. (2021). *Introduction to the Physics and Techniques of Remote Sensing*. Wiley, Hoboken, 3rd ed edition. OCLC: 1243550606.

Ermakov, D., Kuzmin, A., Pashinov, E., Sterlyadkin, V., Chernushich, A., and Sharkov, E. (2021). Comparison of Vertically Integrated Fluxes of Atmospheric Water Vapor According to Satellite Radiothermovision, Radiosondes, and Reanalysis. *Remote Sensing*, 13(9):1639.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B. (2017). The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *Journal of Climate*, 30(14):5419–5454.

Global Modeling And Assimilation Office (2015). M2T3NVASM. https://disc.gsfc.nasa.gov/datacollection/M2T3NVASM_5.12.4.html.

Güsten, R., Nyman, L. A., Schilke, P., Menten, K., Cesarsky, C., and Booth, R. (2006). The Atacama Pathfinder EXperiment (APEX) – a new submillimeter facility for southern skies –. *Astronomy & Astrophysics*, 454(2):L13–L16.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater,

J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049.

Huber, P. J. (1964). Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*, 35(1):73–101.

Iserlohe, C., Fischer, C., Vacca, W. D., Fischer, N., Colditz, S., and Krabbe, A. (2021). Probing the Atmospheric Precipitable Water Vapor with SOFIA. II. Atmospheric Models from ECMWF. *Publications of the Astronomical Society of the Pacific*, 133(1023):055002.

Janssen, M. A., editor (1993). *Atmospheric remote sensing by microwave radiometry*. Wiley series in remote sensing. Wiley, New York.

Johnson, M. D., Akiyama, K., Blackburn, L., Bouman, K. L., Broderick, A. E., Cardoso, V., Fender, R. P., Fromm, C. M., Galison, P., Gómez, J. L., Haggard, D., Lister, M. L., Lobanov, A. P., Markoff, S., Narayan, R., Natarajan, P., Nichols, T., Pesce, D. W., Younsi, Z., Chael, A., Chatterjee, K., Chaves, R., Doboszewski, J., Dodson, R., Doeleman, S. S., Elder, J., Fitzpatrick, G., Haworth, K., Houston, J., Issaoun, S., Kovalev, Y. Y., Levis, A., Lico, R., Marcoci, A., Martens, N. C. M., Nagar, N. M., Oppenheimer, A., Palumbo, D. C. M., Ricarte, A., Rioja, M. J., Roelofs, F., Thresher, A. C., Tiede, P., Weintroub, J., and Wielgus, M. (2023). Key Science Goals for the Next-Generation Event Horizon Telescope. *Galaxies*, 11(3):61.

Kalnay, E. (2003). *Atmospheric modeling, data assimilation, and predictability*. Cambridge University Press, Cambridge. OCLC: 57419695.

Kerber, F. and Querel, R. (2014). All-sky homogeneity of precipitable water vapour over Paranal. *Proceedings of the SPIE*, page 914792.

Kerber, F., Querel, R. R., Rondanelli, R., Hanuschik, R., van den Ancker, M., Cuevas, O., Smette, A., Smoker, J., Rose, T., and Czekala, H. (2014). An episode of extremely low precipitable water vapour over Paranal observatory. *Monthly Notices of the Royal Astronomical Society*, 439(1):247–255.

Kuo, C.-L. (2017). Assessments of Ali, Dome A, and Summit Camp for mm-wave Observations Using MERRA-2 Reanalysis. *The Astrophysical Journal*, 848(1):64.

Li, M.-S., Li, R., Wang, N., and Zheng, X.-W. (2020). Research and application of reanalysis data for radio astronomical site testing. *Research in Astronomy and Astrophysics*, 20(12):200.

Liou, K.-N. (2002). *An introduction to atmospheric radiation.* Number v. 84 in International geophysics series. Academic Press, Amsterdam ; Boston, 2nd ed edition.

Marín, J. C., Pozo, D., and Curé, M. (2015). Estimating and forecasting the precipitable water vapor from GOES satellite data at high altitude sites. *Astronomy & Astrophysics*, 573:A41.

Matsushita, S., Asada, K., Martin-Cocher, P. L., Chen, M.-T., Ho, P. T. P., Inoue, M., Koch, P. M., Paine, S. N., and Turner, D. D. (2017). 3.5 Year Monitoring of 225 GHz Opacity at the Summit of Greenland. *Publications of the Astronomical Society of the Pacific*, 129(972):025001.

Mo, Z., Zeng, Z., Huang, L., Liu, L., Huang, L., Zhou, L., Ren, C., and He, H. (2021). Investigation of Antarctic Precipitable Water Vapor Variability and Trend from 18 Year (2001 to 2018) Data of Four Reanalyses Based on Radiosonde and GNSS Observations. *Remote Sensing*, 13(19):3901.

Otarola, A., De Breuck, C., Travouillon, T., Matsushita, S., Nyman, L.-A., Wootten, A., Radford, S. J. E., Sarazin, M., Kerber, F., and Pérez-Beaupuits, J. P. (2019). Precipitable Water Vapor, Temperature, and Wind Statistics At Sites Suitable for mm and Submm Wavelength Astronomy in Northern Chile. *Publications of the Astronomical Society of the Pacific*, 131(998):045001.

Otung, I. (2021). *Communication engineering principles.* John Wiley & Sons, Inc., USA, Hoboken, NJ, 2nd edition edition.

Paine, S. (2022). The am atmospheric model. *Zenodo.* Publisher: Zenodo Version Number: 12.2.

Radford, S. J. and Holdaway, M. A. (1998). Atmospheric conditions at a site for submillimeter-wavelength astronomy. In Phillips, T. G., editor, *Astronomical Telescopes & Instrumentation*, pages 486–494, Kona, HI.

Raymond, A. W., Palumbo, D., Paine, S. N., Blackburn, L., Córdova Rosado, R., Doeleman, S. S., Farah, J. R., Johnson, M. D., Roelofs, F., Tilanus, R. P. J., and Weintroub, J. (2021). Evaluation of New Submillimeter VLBI Sites for the Event Horizon Telescope. *The Astrophysical Journal Supplement Series*, 253(1):5.

Roelofs, F., Blackburn, L., Lindahl, G., Doeleman, S. S., Johnson, M. D., Arras, P., Chatterjee, K., Emami, R., Fromm, C., Fuentes, A., Knollmüller, J., Kosogorov, N., Müller, H., Patel, N., Raymond, A., Tiede, P., Traianou, E., and Vega, J. (2023). The ngEHT Analysis Challenges. *Galaxies*, 11(1):12.

Shikhovtsev, A. Y., Kovadlo, P. G., Khaikin, V. B., and Kiselev, A. V. (2022). Precipitable Water Vapor and Fractional Clear Sky Statistics within the Big Telescope Alt-Azimuthal Region. *Remote Sensing*, 14(24):6221.

Smette, A., Sana, H., Noll, S., Horst, H., Kausch, W., Kimeswenger, S., Barden, M., Szyszka, C., Jones, A. M., Gallenne, A., Vinther, J., Ballester, P., and

Taylor, J. (2015). Molecfit: A general tool for telluric absorption correction: I. Method and application to ESO instruments. *Astronomy & Astrophysics*, 576:A77.

Ssenyunzi, R. C., Oruru, B., D'ujanga, F. M., Realini, E., Barindelli, S., Tagliaferro, G., von Engeln, A., and van de Giesen, N. (2020). Performance of ERA5 data in retrieving Precipitable Water Vapour over East African tropical region. *Advances in Space Research*, 65(8):1877–1893.

Starr, V. P. and White, R. M. (1955). Direct measurement of the hemispheric poleward flux of water vapor. *Journal of Marine Research*, pages 217–225.

The Event Horizon Telescope Collaboration (2019). First M87 Event Horizon Telescope Results. II. Array and Instrumentation. *The Astrophysical Journal Letters*, page 28.

Tremblin, P., Schneider, N., Minier, V., Durand, G. A., and Urban, J. (2012). Worldwide site comparison for submillimetre astronomy. *Astronomy & Astrophysics*, 548:A65.

Wallace, J. M. and Hobbs, P. V. (2006). *Atmospheric science: an introductory survey*. Number v. 92 in International geophysics series. Elsevier Academic Press, Amsterdam ; Boston, 2nd ed edition. OCLC: ocm62421169.

Xu, J., Li, M., Esamdin, A., Wang, N., Pu, G., Wang, L., Feng, G., Zhang, X., Ma, S., Lv, J., and Zheng, X. (2022). Site-testing at Muztagh-ata site. IV. Precipitable Water Vapor. *Publications of the Astronomical Society of the Pacific*, 134(1031):015006.

Zhao, Y., Yang, F., Chen, X., Zhang, X., Ma, J., Kong, X., Fu, X., Li, R., Wei, Y., Yao, Z., He, F., Pan, J., and Deng, L. (2022). Long-term variations in precipitable water vapor and temperature at Lenghu Site. *Astronomy & Astrophysics*, 663:A34.

Zhu, D., Zhang, K., Yang, L., Wu, S., and Li, L. (2021). Evaluation and Calibration of MODIS Near-Infrared Precipitable Water Vapor over China Using GNSS Observations and ERA-5 Reanalysis Dataset. *Remote Sensing*, 13(14):2761.

Özdemir, S., Yeşilyaprak, C., Aktuğ, B., Öztürk, D., Çoker, D., and Balbay, R. (2018). Precipitable water vapor (PWV) estimations from the site of the Eastern Anatolia Observator (DAG), a new astronomical observatory in Turkey. *Experimental Astronomy*, 46(2):323–336.