Performance Characterization of ESA’s Tropospheric Delay Calibration System for Advanced Radio Science Experiments

Riccardo Lasagni Manghi1, Marco Zannoni1,2, Paolo Tortora1,2, Antonio Martellucci1, Javier De Vicente4, Jose Villalvilla4, Mattia Mercolino4, Gerrit Maschwitz5, and Thomas Rose5

1Dipartimento di Ingegneria Industriale, Alma Mater Studiorum - Università di Bologna, Forlì, Italy, 2Centro Interdipartimentale di Ricerca Industriale Aerospaziale, Alma Mater Studiorum - Università di Bologna, Forlì, Italy, 3European Space Agency, ESA-ESTEC, Noordwijk, Netherlands, 4European Space Agency, ESA-ESOC, Darmstadt, Germany, 5RPG Radiometer Physics GmbH, Meckenheim, Germany

Abstract Media propagation noises are amongst the main error sources of radiometric observables for deep space missions, with fluctuations of the tropospheric excess path length representing a relevant contributor to the Doppler noise budget. Microwave radiometers currently represent the most accurate instruments for the estimation of the tropospheric delay and delay rate along a slant direction. A prototype of a tropospheric delay calibration system (TDCS), using a 14 channel K/V band microwave radiometer, has been developed under a European Space Agency contract and installed at the deep space ground station in Malargüe, Argentina, in February 2019. After its commissioning, the TDCS has been involved in an extensive testbed campaign by recording a total of 44 tracking passes of the Gaia spacecraft, which were used to perform an orbit determination analysis. This work presents the first statistical characterization of the end-to-end performance of the TDCS prototype in an operational scenario. The results show that using TDCS-based calibrations instead of the standard GNSS-based calibrations leads to a significant reduction of the residual Doppler noise and instability.

1. Introduction

Precise radiometric tracking is of key importance during operations of interplanetary missions and for advanced radio science applications. Radio science research performed on deep space missions like Cassini, Juno, BepiColombo, and the upcoming JUICE mission, rely on a combination of X and K band radio links to mitigate the dispersive effects of propagation through interplanetary plasma, solar corona, and Earth ionosphere, leaving tropospheric delay as one of the main error contributors to Doppler and ranging measurements.

To meet the demanding requirements of BepiColombo’s Mercury Orbiter Radio science Experiment (MORE) (Iess et al., 2021) and JUICE’s Geodesy and Geophysics of Jupiter and the Galilean Moons (3 GM) radio science experiment (Cappuccio et al., 2020), in terms of radiometric tracking accuracy and end-to-end stability of the Doppler signals, ground-based microwave radiometers (MWR) are deemed as the most appropriate instruments for tropospheric delay calibration (Iess et al., 2009).

Microwave radiometers have been extensively used in the past decades for the correction of tropospheric induced delay in astrometric and geodetic observations using very large base interferometry (VLBI) (Nikolic et al., 2013; Roy et al., 2007). However, their application to deep space radiometric tracking was originally introduced by the Jet Propulsion Laboratory (JPL) to support the Cassini radio science experiments, with the development and installation of two dedicated high-stability water vapor radiometers at the Goldstone deep space communication complex in California (Naudet et al., 2000; Resch et al., 2001; Tanner & Riley, 2003).

In the following years, in combination with multi-frequency links to calibrate the dispersive media, such as the ones available on Cassini and Juno, the use of these radiometers allowed to reach a remarkable end-to-end fractional frequency stability, expressed in terms of Allan standard deviation, of about $1 \times 10^{-14}$ at a 1,000 s stability interval (Durante et al., 2020; Tortora et al., 2004).
Later, the European Space Agency (ESA) performed a preliminary study named AWARDS (Graziani, et al., 2014; Tortora, et al., 2013) for the definition of the requirements and preliminary system design of a tropospheric delay calibration system (TDCS). In addition, media calibration performance requirements for accurate spacecraft tracking were studied in detail in another ESA study called ASTRA (Iess et al., 2012, 2014).

The TDCS described here represents the prototype of a new instrument for the calibration of tropospheric delay and delay rate based on a high stability and high accuracy K/V band MWR, which was developed in the framework of an ESA contract by a consortium formed by Radiometer Physics GmbH (RPG), University of Bologna, and the Université Catholique de Louvain.

Specifically, this work focuses on the end-to-end performance characterization of the TDCS, which was carried out through a detailed orbit determination using Doppler measurements acquired from ESA’s deep space ground station (DSA3) in Malargüe, during several tracking passes of the Gaia spacecraft.

It is important to clarify that the purpose of this test is not to reproduce the full orbit determination solution used for the navigation of Gaia but to validate the TDCS products by making a punctual evaluation of the relative end-to-end noise reduction when TDCS-based calibrations are used in place of standard calibrations based on global navigation satellite system (GNSS) measurements.

2. Tropospheric Delay Calibration System

The TDCS is a combination of instruments, software tools, and operational procedures that allows an accurate estimation of the tropospheric delay along the slant direction while minimizing the effect of the instrument instability. The main TDCS subsystem is an ultra-stable MWR for deep-space applications, which represents a modified version of the standard HATPRO-G5 radiometer developed by RPG (Maschwitz et al., 2019). A description of the concept and design of the first HATPRO generation is given by Rose et al. (2005). In 2015, the radiometric performance was significantly improved with the fifth generation of the series (G5).

The TDCS, shown in Figure 1, measures sky noise emissions at frequencies near the water vapor absorption peak at 22.2 GHz, the oxygen absorption band around 60 GHz, and in the 30 GHz window, which is sensitive to liquid water content (clouds and rain). With respect to the standard HATPRO-G5 radiometer, this instrument was tailored for spacecraft tracking applications by adding (a) a 2-axes tracking system to gain full sky scanning capabilities, (b) a modified antenna structure including an external heated parabolic reflector, which narrows the MWR half-power beamwidth down to 1.2°, allowing to better replicate the air volume sampled by the deep space antenna and to reduce the effect of solar radiation contamination during periods of superior conjunction, and (c) a high precision meteo station providing values of air pressure, temperature, relative humidity, rain rate and wind speed at ground level. The instrument includes internal and external temperature control and antenna blower systems to avoid water condensation and the icing on the exposed surfaces and to maintain accurate stability of the receivers’ temperature.

Further updates include new control procedures for the tracking system, specific calibration procedures for the instrument parameters, and the development of an external software tool that acts as a high-level access point for monitoring and controlling the TDCS functions by the ground station front-end controller.

The TDCS was installed at ESA’s deep space ground station in Malargüe in February 2019 and is located at a distance of 31.8 m from the deep space antenna along the North-North-East direction. This physical separation represents a relevant error source for the delay and delay-rate calibrations, due to the different air volumes sampled by the radiometer and the deep space antenna. This observation prompted JPL to propose a new design for its next generation of a tropospheric calibration system, which foresees a direct integration...
of the water vapor radiometer within the antenna receiver (Tanner et al., 2021). While a direct integration was not possible for the current setup, the selected mounting location for the TDCS is well within the 50 m distance limit proposed by Tortora et al. (2013) and represents a tradeoff between calibration performances and operational constraints (i.e., availability of supporting infrastructures and visibility during tracking).

3. Testbed Summary and Data Availability

Two successive testbed campaigns were carried out between February and July 2019, targeting a total of 44 tracking passes of ESA’s Gaia spacecraft. Both hardware and software updates were performed in response to issues that were encountered during the test campaign. As a result, the availability of key data products (Doppler data, GNSS-based calibrations, and TDCS-based calibrations) has varied among the different tracking passes. Since the simultaneous availability of all these products is required for the performance evaluation, only 32 of the total passes were eventually included within the analysis, as shown in Table 1, which provides an overview of this subset along with a summary of the atmospheric conditions that were encountered during each pass.

The range of TDCS-retrieved zenith wet delay (ZWD) values indicates the potential improvement that can be obtained when tropospheric calibrations are introduced in the orbit determination process. Being Gaia only visible at night, most of the tracking passes are characterized by dry conditions, which can limit the effectiveness of the TDCS calibrations. Values of TDCS-retrieved liquid water path (LWP) above \( \sim 10 \, \text{g/m}^2 \) indicate the presence of condensed water (clouds or fog) along the instrument line-of-sight. These values scale with the length of the propagation path through the cloud, so high values (i.e., \( \sim 500–1,000 \, \text{g/m}^2 \)) indicate the presence of thick cloud formations and may suggest the presence of precipitated water, in particular when coupled with the triggering of the rain flag (RF) by the rain sensor included within the TDCS. Characteristic integrated values for wind speed (WS) and turbulence strength \( (C_{n2}) \), derived from the European Center for Medium Weather Forecast (ECMWF) database (Molteni et al., 1996), can be considered as proxy parameters for the presence of turbulent eddies in the lower portions of the atmosphere, which affect the accuracy of the TDCS calibrations (Lasagni Manghi et al., 2019). Both values of wind speed and turbulence strength were provided by UCL (Quibus et al., 2019) and were derived by averaging over time the local vertical profiles from the ECMWF data set that fell within the tracking pass interval. The obtained vertical profiles were then spatially averaged from ground level to a height of 1 km above the surface.

Finally, the wind speed at ground level, measured by the TDCS meteo station, provides, on one hand, an indication of the strength of the mechanical noise introduced in the Doppler measurements by the tropospheric calibrations, and on the other hand, represents another proxy for the effect of wind shears on tropospheric turbulence.

4. TDCS Data Processing

The retrieval of atmospheric variables from MWR observations is an ill-posed problem, since a given set of brightness temperature measurements may be related to several different atmospheric states (Keihm & Marsh, 1998). To resolve this ambiguity, the TDCS uses a neural network retrieval algorithm, which was trained using a large number of atmospheric vertical profiles extracted specifically for the Malargüe site from a numerical weather prediction model (ECMWF reanalysis). From each of these profiles of temperature, humidity, pressure, and liquid water concentration, the slant wet delay was computed. At the same time, each profile was used as input for the simulation of the corresponding brightness temperature measurements via a state-of-the-art radiative transfer model. The resulting data set was split into a training part and a test part. For the former, simulated brightness temperatures and slant wet delay values were used to derive a set of retrieval coefficients, which minimized the output error over the complete training data set. The latter served for validation of the retrieval performance through a statistical comparison of the slant wet delay values calculated from the test profiles and the ones retrieved via the neural network coefficients.

During operations, the retrieval coefficients related the measured brightness temperature vector (14 K/V band channels) to the best slant wet delay value. Since brightness temperature and delay scale differently with the length of the propagation path through the atmosphere, the estimation process of the retrieval
parameters was repeated at 19 discrete elevation angles (spaced at constant airmass steps) covering the range between 10° and 90°, leading to the generation of an elevation-dependent grid of retrieval coefficients. The discrete angle-grid may cause artificial effects at low elevation angles (<30°). This was handled by a smart interpolation scheme based on the following steps: at first, the mean radiative temperature was derived from surface sensors observations and a dedicated neural network retrieval and used to derive the atmospheric opacity from brightness temperatures. Then, the atmospheric opacity was linearly interpolated to the nearest nodes on the retrieval grid and used to derive the corresponding brightness temperature.
values using the mean radiative temperature and neglecting its small variation with airmass. Finally, the slant wet delay output was calculated as an airmass weighted average of the values retrieved for the two considered grid nodes.

For each of the analyzed tracking passes, slant wet delay values were retrieved from brightness temperature measurements using the trained neural network retrieval and converted to zenith wet delay using a $\sin(\theta)$ mapping function, where $\theta$ represents the instantaneous elevation as indicated by the tracking system. This simple mapping was preferred to higher fidelity mapping functions, such as the one from Niell (1996), to be consistent with the procedures used for the retrieval training.

Then, outliers were identified by computing the distance of each data point from a smoothed data set, which was generated using a median filter with a 10 min time window, and removed when lying outside a fixed number of standard deviations from the median.

Finally, calibration cards were generated from the zenith wet delay time series using a linear piecewise fit between consecutive data points and according to the definitions provided by JPL (Media Calibration Interface, 2008). Several calibration cards were generated with increasing values of the zenith wet delay integration time to improve the signal-to-noise ratio. Specifically, TDCS calibrations with 1, 20, and 60 s integration time were respectively used and compared within the orbit determination analysis.

The zenith hydrostatic delay (ZHD) was computed according to the model of Saastamoinen (1972), using Equation 1. In this expression, $p_s$ (mbar) is the surface pressure measured by the TDCS meteo sensor, $\lambda$ (°) is the geocentric latitude, and $h$ (m) is the height over the mean sea level. Geographic coordinates for the TDCS, which were used in Equation 1, are provided in Table 2.

$$ZHD = 2.2767 \times 10^{-3} \frac{p_s}{1 - 0.0266 \times \cos(2\lambda) - 0.00028 \times h} (m)$$

The ZHD was then scaled to the height of the deep space antenna using the correction given in Equation 2, where $\Delta h$ (m) is the height difference between the TDCS and the antenna reference point (corresponding to the azimuth and elevation axes crossing), $p_s$ (mbar) is the surface pressure, and $T_s$ (K) is the surface temperature. This expression is a modified version of the one from Estefan and Sovers (1994), where the average pressure and temperature of the vertical air column were replaced by the instantaneous surface measurements provided by the TDCS meteo station.

$$\Delta ZHD = -7.76 \times 10^{-3} \frac{\Delta h}{T_s} (m)$$

The corrected delay values were then smoothed using a Gaussian filter with a 10 min time window and down-sampled to 20 s to remove some discontinuities within the 1 Hz sampling rate time series. These discontinuities, which were caused by the resolution of the TDCS pressure sensor of 0.1 mbar (corresponding to a zenith delay resolution of about 0.23 mm), were comparable in magnitude to the short scale variations in the zenith wet delay and might have resulted in an increased Doppler noise if not properly corrected. Finally, calibration cards were generated from the ZHD time series using a linear piecewise fit between consecutive data points.

### 5. Orbit Determination Analysis

#### 5.1. Introduction

Gaia is an ESA cornerstone scientific mission, whose aim is to measure the three-dimensional position and velocity distributions of stars within the Milky Way using accurate astrometric measurements (Prusti et al., 2016). The selection of this particular mission for the TDCS testbed campaign was mainly driven by geometrical and operational considerations. Since Gaia operates from a Lissajous-type orbit around the second Earth-Sun Lagrange point, it is constantly near solar opposition. This means that the impact of solar plasma and Earth ionosphere on the propagation delay is particularly limited, thus simplifying the processing and calibration procedures for the Doppler data. Furthermore, several Gaia tracking passes were already...
scheduled at the DSA3 complex during the same time period, thus the inclusion of TDCS operations had a marginal impact on ground station operations.

The overall concept for this analysis was to perform a standard orbit determination process for the Gaia spacecraft, using Doppler measurements collected at the DSA3 complex and *a priori* information on the dynamical model provided by the Flight Dynamics team at ESA’s European Space Operation Center in Darmstadt, Germany (ESOC FD). This process was repeated by keeping all parameters fixed while varying the applied tropospheric calibrations (either generated from dual-frequency GNSS measurements or generated by the TDCS measurements) to allow for a direct comparison of the respective accuracies.

5.2. Data Selection and Processing

Raw Doppler measurements at X/X band acquired during the Gaia tracking passes between February 16 and July 18, 2019 were provided by ESA as collected by the telemetry, tracking and command processor, according to the format definitions provided by Ricart (2018). As a first step, the set of Doppler observables was reduced by removing all measurements in the proximity of the chemically propelled maneuvers to avoid discontinuities.

Then, all observables collected below 15° of elevation at the ground station were removed to mitigate the progressive degradation of the radiometric retrieval accuracy. This value, which represents a conservative limit, was selected to account for the retrieval errors due to the granularity of the retrieval coefficients, possible fast variations of the observed atmospheric scene, and contaminations due to the ground and clutter emission.

Doppler measurement weights for each tracking pass were computed as the root mean square value of the residuals for that pass.

5.3. Media Calibrations

For the most recent deep space missions, the dispersive effect from the charged particles in the solar corona is calibrated using a multi-frequency link with coherent uplink and downlink (Bertotti et al., 1993; Mariotti & Tortora, 2013). This was not possible for the current analysis, since Gaia uses a single frequency link at X-band. However, the effect of Solar plasma is assumed to be small, considering that Gaia operates near solar opposition, with elongation values always larger than 170° (Asmar et al., 2005; Iess et al., 2014).

Similar considerations apply for the propagation delay induced by charged particles in the Earth’s ionosphere. Being Gaia only visible at night, the ionospheric induced Doppler error at X-band was expected to be small when compared to the variations of tropospheric delay (Thornton & Border, 2003), with typical values in the order of 10 μm/s, and was therefore neglected.

The GNSS-based tropospheric calibrations, which were used as a baseline for validation of the TDCS products, were provided by ESOC FD in the form of time-normalized polynomials of 6 h intervals, according to the format described by JPL (Media Calibration Interface, 2008). The tropospheric zenith delay was estimated from GNSS measurements with a 5 min temporal resolution, using the precise point positioning Kalman filter approach (Zumberge et al., 1997). The ZHD was then derived from surface pressure measurements, following the same procedure described in Section 4, and subtracted from the total delay to obtain the wet component (Feltens et al., 2018).

5.4. Dynamical Model

The dynamical model was kept reasonably simple to reduce the likelihood of possible biases in the results caused by mismodelling errors.

The gravitational accelerations that were considered for this analysis include point-mass gravity from the Sun, the planets and their satellites, the Moon, and Pluto. Higher order gravitational harmonics were neglected. State vectors and gravitational parameters for the Solar System bodies were taken from JPL’s DE430 planetary ephemerides (Folkner et al., 2014).
Non-gravitational accelerations were introduced in the form of interpolating polynomials using tabulated coefficients generated by ESOC FD, allowing for easier replicability. Specifically, the main non-gravitational accelerations acting on the spacecraft were the ones from solar and thermal radiation pressure, which were provided in form of normalized acceleration components.

Gaia performed three main orbit trim maneuvers (OTM) throughout the testbed campaign: two station-keeping maneuvers, in February and April 2019, respectively, and an inclination change maneuver, split into nine burns, in July 2019. All maneuvers occurring during the tracking passes were modeled as impulsive burns and estimated within the filter using a priori ΔV values provided by ESOC FD.

Attitude control during chemically propelled maneuvers was performed using the reaction control system (RCS), which caused parasitic ΔVs to be imparted on the spacecraft. Similar to the OTM, reaction control system firings were modeled as impulsive burns and estimated within the filter.

Conversely, attitude control during standard operations was performed using a cold-gas micro-propulsion system, which caused parasitic accelerations to act permanently on the spacecraft. Instantaneous accelerations from cold-gas thrusters were provided in the tabulated form by ESOC FD.

### 5.5. Filter Setup

The analysis was carried out using JPL’s MONTE orbit determination software (Evans et al., 2018), which adopts a weighted least-squares batch filter to generate iterative corrections to the a priori dynamical model in order to minimize the difference between the real and the simulated measurements (Bierman, 2006).

Table 3 summarizes the solve-for parameters within the square root information batch filter and their associated a priori uncertainties.

| Parameter                  | Type  | N_{est} | A priori σ | Comments                                             |
|----------------------------|-------|---------|------------|------------------------------------------------------|
| Spacecraft position        | Local | 3       | 100 km     | A priori values were taken from the ESOC FD solution.|
| Spacecraft velocity        | Local | 3       | 1 m/s      |                                                      |
| Virtual antenna position   | Local | 2       | 10 cm      | A priori coordinates in the spacecraft body frame     |
| ΔV (OTM)                   | Global| 3 × N_{OTM} | 10 cm/s   | A priori values were provided by ESOC FD            |
| ΔV (RCS)                   | Global| 3 × N_{RCS} | 1 cm/s    |                                                      |

A priori values for the spacecraft state were taken from the operational trajectory reconstructed by ESOC FD. Another key parameter that was estimated within the filter is the phase center of the onboard antenna. Since Gaia uses two separate antennas for uplink and downlink, which are respectively a low-gain antenna and a phased array antenna, the estimated coordinates are actually referred to a virtual antenna located at the midpoint between the two. Estimated values for the coordinates of the virtual antenna were consistent with the a priori uncertainty and mostly absorbed the short-term variations in the location of the spacecraft center of mass. It should be noted that the virtual antenna coordinate along the spin axis, which corresponds to the x-axis of the spacecraft body frame, is not observable using Doppler measurements, so only the y and z components were estimated locally for each pass.

### 6. Results

The overall quality of the TDCS calibrations is driven by several intrinsic and scene-dependent factors. The former comprises all error sources which are related to the MWR components, such as the noise characteristics of the MWR radiometric receivers, antenna losses, and spill-over losses of the K band channels over a variable background. The latter comprises all error sources whose magnitude depends on the local atmospheric conditions encountered during the measurements. These include the retrieval error contribution,
which depends on the completeness and variety of the training database, and a beam mismatch contribution related to the different air volumes observed by the TDCS and the deep space antenna.

When the tropospheric calibrations are included within the orbit determination process, the radiometric measurements will thus be affected by a variable amount of uncalibrated (or residual) tropospheric delay and by additional error sources that are introduced as part of the calibration generation process.

The slant wet delay retrieval accuracy of G5 radiometers was estimated through a series of retrieval self-tests using portions of the ECMWF training database for Malargüe. Its value ranged between 0.8 mm at Zenith and 6.3 mm at 10° of elevation, well below the 20 mm accuracy requirement from the AWARDS study (Tortora, et al., 2013). These accuracies should be applicable to the TDCS as well, under the assumption that possible antenna losses are corrected effectively. However, an end-to-end assessment of the delay accuracy using range observations was harder to obtain, due to the simplified dynamical model employed for this analysis.

The key factor when assessing the impact of tropospheric calibrations on Doppler measurements is represented by the calibration stability, which is usually characterized in terms of Allan standard deviation at various time intervals. The stabilities of the radiometric receivers (instrumental noise) and of the retrieval algorithm were the subject of previous investigations by the authors (Lasagni Manghi, et al., 2019; Maschwitz et al., 2019), and were numerically quantified with simulations and testing in controlled environments. The former showed typical values of Allan deviation of $1.3 \times 10^{-15}$ at 1,000 s, while the latter ranged between $\sim 10^{-16}$ and $\sim 10^{-15}$ at 1,000 s for the case of “broken cloud advection” (alternating sequences of clear-sky and cloudy conditions), which was deemed as the worst-case retrieval scenario. Analogously, the beam mismatch contribution was quantified by numerical simulations under specific atmospheric conditions (Graziani, et al., 2014).

Instead of focusing on individual error contributions, the current analysis provides an estimation of the end-to-end frequency stability of the Doppler residuals obtained when using either GNSS-based or TDCS-based tropospheric calibrations.

As a first step, a visual inspection of the Doppler residuals at 60 s count time was performed to highlight the presence of major signatures within the data and to identify possible causes for these features. The specific value of 60 s for the count time was selected since it represents a standard case for radio science applications (Casajus et al., 2021; Durante et al., 2019; Tortora et al., 2016; Zannoni et al., 2020). In fact, this value is sufficiently smaller than the characteristic time scales of the typical investigated processes and sufficiently large to avoid numerical noise issues (Zannoni & Tortora, 2013). Root mean square (RMS) values of the residuals were then produced for each pass, along with an estimation of the relative noise reduction between the two analyzed cases.

Finally, the overall stability of the Doppler residuals was quantified by computing the Allan standard deviation (ASD) according to Equation 3, where $y$ represents the normalized frequency residuals, $\Delta T$ is the stability time interval, the brackets $\langle \cdot \rangle$ indicate an ensemble average over the measured time series, and $\langle y(t,\Delta T) \rangle$ indicates a time average over the interval between $t$ and $t + \Delta T$, according to the expression in Equation 4. Specifically, stability intervals of 20, 60, and 1,000 s were considered, which represent typical values used for radio science applications.

$$\text{ASD}_y(\Delta T) = \left( \frac{1}{2} \langle y(t + \Delta T,\Delta T) - y(t,\Delta T) \rangle^2 \right)^{1/2}$$

$$y(t,\Delta T) = \frac{1}{\Delta T} \int_{t}^{t+\Delta T} y(\tau) d\tau$$

In the following, a single representative pass is analyzed in detail according to the procedure described above. Then, a summary of all passes is produced with a quantitative comparison of the relative performances between the analyzed test cases.
6.1. Example Tracking Pass (April 19, 2019)

This pass was selected as representative of standard conditions that were encountered at the deep space ground station in Malargüe. From Table 1, we observe that this pass was characterized by moderate liquid water content and that the rain flag was not triggered, suggesting the presence of clouds along the line of sight, with no precipitation. Moderate to high values of wind speed were also observed at ground level, in particular during the first hours of the pass. Vertical profiles from the ECMWF data set also suggest the presence of moderate to high turbulence levels.

Figure 2 compares the Doppler residuals at 60 s count time, using GNSS-based calibrations (left) and TDCS-based calibrations, with 20 s integration time (right). With the introduction of TDCS calibrations, we observe a consistent improvement in the residuals, with an overall 51% reduction of the root mean square values and no apparent signature being introduced. The observed improvement is particularly pronounced during the first half of the tracking pass, where higher wind speed and liquid water path values are observed. This may provide an indication of the ability of the selected neural network retrieval to correctly separate the information content of liquid water from the one provided by water vapor.

The left plot in Figure 3 shows a comparison of the ASD curves obtained by applying Equation 3 to the Doppler residuals for four analyzed test cases, corresponding to GNSS calibrations, and TDCS calibrations at 1, 20, and 60 s integration time, respectively. Till stability intervals of 10 s, all curves, with the exception of the TDCS calibrations at 1 s integration time, are collapsed and approximately follow a power law with a slope equal to −1. This behavior may suggest that the dominant error source at those characteristic timescales is the Doppler thermal noise (Jess et al., 2014). However, for short integration times of the tropospheric products, the thermal noise of the MWR receiver components, which is introduced through the calibrations, becomes comparable in magnitude and induces the observed offset in the 1 s curve.

At longer stability intervals, the uncalibrated tropospheric delay becomes progressively more relevant, as indicated by the departure of all curves from the initial linear trend. From Figure 3, it is clear that the TDCS-based calibrations are able to capture the atmospheric variability along the slant path much better than their GNSS-based counterpart, with minimum ASD values that are obtained for a 20 s integration time of the zenith wet delay, which is therefore used in the following sections for the overall performance characterization.
Similar results are observed through a comparison of the power spectral density of the Doppler residuals, which were generated using an adaptive multi-taper spectral estimation method (Percival & Walden, 1993). From the right plot of Figure 3 we can observe that above $10^{-1}$ Hz the spectrum is dominated by the Doppler thermal noise. Most of the atmospheric instability that is calibrated using TDCS data occurs at characteristic frequencies between $10^{-4}$ and $10^{-2}$ Hz.

### 6.2. Overall Statistics

The procedure described above was repeated for all the passes included within the analysis. Figure 4 compares the Doppler residuals at 60 s count time, using GNSS-based calibrations and TDCS-based calibration, respectively, for the whole testbed campaign. Even though the root mean square value is dominated by the presence of a few noisy passes, likely caused by adverse meteorological conditions, an overall improvement of the residuals is clearly detectable.

Figure 5 shows the root mean square values for the Doppler residuals as a function of the pass ID, along with the ratio between corresponding values for the two test cases. The average noise reduction between the different tracking passes is approximately 34% when using TDCS-based calibrations instead of the GNSS-based ones, with a maximum reduction of 61% for pass 32 (July 18, 2019). Although with different magnitudes, all passes show a noise reduction, with the exception of pass 31 (July 16, 2019) for which the noise increased by approximately 11%. However, this pass incidentally coincided with a series of reaction control system firings and OTM, which increased the number of estimated parameters and limited the availability of Doppler observables. Moreover, by looking at Table 1 we can observe that this pass was characterized by extremely dry conditions, which corresponded to a reduced signal-to-noise ratio of the estimated calibrations.

Considering the limited number of observed tracking passes, it is difficult to pinpoint an exact cause for the variability in the performance of the TDCS products. The amount of uncalibrated atmospheric variability affecting the Doppler residuals depends both on the actual value of the integrated zenith wet delay and on the accuracy of the calibrations, which strongly depends on the atmospheric conditions. Using TDCS calibrations may also introduce additional error sources such as the mechanical noise from wind-induced vibrations of the tracking system mounting structure or radiometric retrieval errors induced by fast
variations in the observed atmospheric scene (in particular at low elevations), which may dominate over the tropospheric noise for particular tracking passes.

Finally, Figure 6 offers a comparison of the ASD curves computed at characteristic stability intervals of 20 s, 60 s, and 1,000 s, respectively, which represent typical values used for radio science applications. It can be observed that both the 20 and 60 s curves show a consistent reduction of the ASD values when using TDCS-based calibrations, with a magnitude that is more pronounced for the latter case. A similar reduction is observed for the 1,000 s stability interval curves, with the exception of a couple of tracking passes, corresponding to pass IDs 11 and 14. A detailed inspection of the Doppler residuals for these tracking passes highlighted the presence of small wave-like signatures at elevation angles below 30°, which was introduced by the TDCS calibrations. The cause of these signatures, which is currently under investigation, is expected

Figure 4. Comparison of Gaia Doppler residuals for the whole testbed campaign. Left: using global navigation satellite system calibrations; right: using tropospheric delay calibration system calibrations with 20 s integration time.

Figure 5. Left: root mean square (RMS) values for the Doppler residuals at 60 s count time using global navigation satellite system calibrations (gray triangles) and tropospheric delay calibration system calibrations with 20 s integration time (black circles); right: ratio between RMS values for the two test cases.
to be related to the granularity of the elevation-dependent retrieval coefficients. The retrieval-induced error, which is small for most of the tracking passes, may become relevant for specific atmospheric conditions and particularly at low elevation angles, for which the observed atmospheric scene may be subject to fast variations.

Additional investigations may be required for a fine-tuning of the retrieval algorithm, which could improve the accuracy of the tropospheric calibrations at low elevations.

7. Conclusions

This work presented the first statistical characterization of the end-to-end performances of the TDCS prototype that was installed at ESA’s deep space ground station in Malargüe, Argentina.

An extensive testbed campaign was carried out between February and September 2019, using the TDCS alongside the 35 m deep space antenna to track the Gaia spacecraft during a series of scheduled passes. The described analysis, which does not replicate the full orbit determination solution for the navigation of Gaia, was mostly intended as a side-by-side comparison of the orbit determination performance when TDCS-based tropospheric calibrations are used in place of the standard GNSS-based calibrations.

The instrument performance was characterized in terms of root mean square values of the Doppler residuals and Allan standard deviation computed at characteristic stability intervals. The results indicate that an average reduction of about 34% in the residual Doppler noise is observed when TDCS-based calibrations are used. The actual magnitude of this improvement strongly varies between the different tracking passes, with maximum reductions around 61% and a few cases with no appreciable improvement. The overall quality of TDCS calibrations depends on several factors, including the magnitude of the actual tropospheric variability (which depends on the integrated water vapor content along the slant direction), the accuracy of the neural network retrieval, and the magnitude of the additional error sources introduced by the calibration process.

A complete statistical characterization of the TDCS performances would require the analysis of a larger sample of tracking passes under diverse observing conditions (e.g., tracking during daytime or at low
el elongation values). Future work may therefore include additional observations for Gaia, along with the analysis of large data sets for BepiColombo, Mars Express, or the ExoMars orbiter, which are routinely being tracked from the DSA3 station complex.

More specifically, an analysis of BepiColombo tracking passes is currently underway as part of the cruise and solar conjunction radio science experiments. This analysis is expected to improve the TDCS performance characterization, thanks to the more accurate $K_a/K_b$ band tracking link that allows for an almost complete cancellation of the solar and ionospheric plasma noises.

The inclusion of ranging data within the orbit determination setup, combined with a high-fidelity dynamical model and longer integration arcs, will also enable an end-to-end assessment of the delay accuracy, which is relevant for possible applications of TDCS calibrations to VLBI measurements.

Data Availability Statement

Data used for this research will be made publicly available through the Guest Storage Facility (GSF) within ESA's Planetary Science Archive (https://www.cosmos.esa.int/web/psa/psa_gs). This data set will include all raw measurements and ancillary information required for replicating the Gaia orbit determination analysis.

Acknowledgments

The research described in this paper was carried out at the University of Bologna in the framework of ESA contract No. 4000116932/16/NL/AF “Development of a Ground Tropospheric Media Calibration System for Accurate Ranging of Space Science Missions.” The authors want to express their gratitude to the personnel involved in the TDCS project who gave precious suggestions for completing this study and in particular Alberto Graziani and Gaullme Autret from the European Space Agency, the ESOC Flight Dynamics team in the persons of Frank Budnik and Gabriele Belli, the GAIA mission team, Emiliano Orlandi and Harald Czerkala from RFG for the retrieval development and tests, Danielle Vanhoeacker-Janvier and Laurent Quibus from Université Catholique de Louvain-la-Neuve, and Gilles Mariotti from SITAEL S.p.A. R. Lasagni Manghi, M. Zannoni, and P. Tortora wish to acknowledge Caltech and the Jet Propulsion Laboratory for granting the University of Bologna a license to an executable version of the CRUI-CARE Agreement.

References

Asmar, S., Armstrong, J., Jess, L., & Tortora, P. (2005). Spacecraft Doppler tracking: Noise budget and accuracy achievable in precision radio science observations. Radio Science, 40(2). https://doi.org/10.1029/2004RS003101
Bertotti, B., Comoretto, G., & Jess, L. (1993). Doppler tracking of spacecraft with multi-frequency links. Astronomy and Astrophysics, 269, 608–616. http://adsabs.harvard.edu/pdf/1993A%26A...269..608B
Bierman, G. J. (2006). Factorization methods for discrete sequential estimation. Courier Corporation.
Cappuccio, P., Hickey, A., Durante, D., Di Benedetto, M., Jess, L., De Marchi, F., et al. (2020). Ganimede’s gravity, tides and rotational state from JUICE’s 3GM experiment simulation. Planetary and Space Science, 187, 104902. https://doi.org/10.1016/j.pss.2020.104902
Casasús, L. G., Zannoni, M., Modenini, D., Tortora, P., Nimmo, F., Van Hoolst, T., et al. (2021). Updated Europa gravity field and interior structure from a reanalysis of Galileo tracking data. Icarus, 358, 114187. https://doi.org/10.1016/j.icarus.2021.114187
Durante, D., Hemingway, D. J., Racioppa, P., Jess, L., & Stevenson, D. J. (2019). Titan’s gravity field and interior structure after Cassini. Icarus, 326, 123–132. https://doi.org/10.1016/j.icarus.2019.03.003
Durante, D., Parisi, M., Serra, D., Zannoni, M., Notaro, V., Racioppa, P., et al. (2020). Jupiter’s gravity field halfway through the Juno mission. Geophysical Research Letters, 47(4), 1–8. https://doi.org/10.1029/2019GL086572
Estefan, I., & Sovers, O. (1994). A comparative survey of current and proposed tropospheric refraction-delay models for DSN radio metric data calibration.
Evans, S., Taber, W., Drain, T., Smith, J., Wu, H.-C., Guevara, M., et al. (2018). MONTE: The next generation of mission design and navigation software. CEAS Space Journal, 10(1), 79–86. https://doi.org/10.1007/s12567-017-0171-7
Feltens, J., Belli, G., Springer, T., Kints, M. V., Zandbergen, R., Budnik, F., & Schönhemann, E. (2018). Tropospheric and ionospheric media calibrations based on global navigation satellite system observation data. Journal of Space Weather and Space Climate, 8, A30. https://doi.org/10.1051/swsc/2018016
Folkner, W., Williams, J., Boggs, D., Park, R., & Kuchynka, P. (2014). The planetary and lunar ephemerides DE430 and DE431. Icarus, 238, 1–29. https://doi.org/10.1016/j.icarus.2014.03.018
Folkner, W., Williams, J., Boggs, D., Park, R., & Kuchynka, P. (2014). The planetary and lunar ephemerides DE430 and DE431. Interplanetary Network Progress Report 196(1). Retrieved from https://ipnpr.jpl.nasa.gov/progress_report/196/196C.pdf
Fornasier, S. V., Fornasier, S. V., & Strömberg, J. (2009). Assessment of a ground-based microwave radiometer calibration. IEEE Transactions on Antennas and Propagation, 67(2), 623–6241. https://doi.org/10.1109/TAP.2013.2307582
Iess, L., Asmar, S., & Tortora, P. (2009). MORE: An advanced tracking experiment for the exploration of Mercury with the mission BepiColombo. Acta Astronautica, 65(5–6), 666–675. https://doi.org/10.1016/j.actaastro.2009.01.049
Iess, L., Asmar, S. W., Cappuccio, P., Cassioli, G., De Marchi, F., di Stefano, I., et al. (2021). Gravity, geodesy and fundamental physics with BepiColombo’s MORE investigation. Space Science Reviews, 217(1), 21. https://doi.org/10.1007/s11214-021-00800-3
Iess, L., Budnik, F., Colamarrino, C., Corbelli, A., Di Benedetto, M., Fabbri, V., et al. (2012). ASTRA: Interdisciplinary study on enhance-
Maschwitz, G., Czekala, H., Orlandi, E., & Rose, T. (2019). Accuracy and performance of atmospheric delay by a RPG microwave radiometer with respect to ground calibration systems for ESA radio science. TT&C workshop.

Molteni, F., Buizza, R., Palmer, T., & Petroliagis, T. (1996). The ECMWF ensemble prediction system: Methodology and validation. Quarterly Journal of the Royal Meteorological Society, 122(529), 73–119. https://doi.org/10.1002/qj.4791225905

Naudet, C., Jacobs, C., Keihm, S., Lanyi, G., Linfield, R., Resch, G., et al. (2000). The media calibration system for Cassini radio science: Part I. The Telecommunications and Mission Operations Progress Report (pp. 42–123). Retrieved from https://50.18.115.94/progress_report/42-143/143I.pdf

Niel, A. (1996). Global mapping functions for the atmosphere delay at radio wavelengths. Journal of Geophysical Research: Solid Earth, 101(B2), 3227–3246. https://doi.org/10.1029/95JB03048

Nikolic, B., Bolton, R. C., Graves, S. F., Hills, R. E., & Richer, J. S. (2013). Phase correction for ALMA with 183 GHz water vapour radiometers. Astronomy & Astrophysics, 552, A104. https://doi.org/10.1051/0004-6361/201220987

Percival, D., & Walden, A. (1993). Spectral analysis for physical applications. Cambridge University Press.

Prusti, T., De Bruijne, J. H. J., Brown, A. G., Vallenari, A., Babusiaux, C., Bailey-Jones, C. A. L., et al. (2016). The gaia mission. Astronomy & Astrophysics, 595, A1. https://doi.org/10.1051/0004-6361/201629272

Quibus, L., Luini, L., Riva, C., & Vanhøenacker-Janvier, D. (2019). Use and accuracy of numerical weather predictions to support EM wave propagation experiments. IEEE Transactions on Antennas and Propagation, 67(No. 8). https://doi.org/10.1109/TAP.2019.2913785

Resch, G. M., Clark, J. E., Keihm, S. J., Lanyi, G. E., Naudet, C. J., Riley, A. L., et al. (2001). Resch, G. M., Clark, J. E., Keihm, S. J., Lanyi, G. E., Naudet, C. J., Riley, A. L., et al. (2001). The media calibration system for Cassini radio science: Part II. TMO Progress Report (pp. 42–145). Retrieved from https://tda.jpl.nasa.gov/progress_report/42-145/1453.pdf

Ricart, M. (2018). TTCP Software Interface Control Document (ICD) for RM datasets (Issue 2.1).

Rose, T., Crewe, S., Löhnter, U., & Simmer, C. (2005). A network suitable microwave radiometer for operational monitoring of the cloudy atmosphere. Atmospheric Research, 75(2005), 183–200. https://doi.org/10.1016/j.atmosres.2004.12.005

Roy, A. L., Rottmann, H., Teuber, U., & Keller, R. (2007). Phase correction of VLBI with water vapour radiometry. arXiv:astro-ph/0703066.

Saastamoinen, J. (1972). Atmospheric correction for the troposphere and stratosphere in radio ranging satellites. In The use of artificial satellites for geodesy (pp. 247–251). American Geophysical Union (AGU). Retrieved from http://doi.wiley.com/10.1029/GM015p0247

Tanner, A., Border, J., Jongeling, A., Long, E., Pereira, E., & Lin, E. (2021). Embedding a water vapor radiometer within a deep space network Ka-band receiver. The Interplanetary Network Progress Report, 42–226, 1–17. https://ipnwww.jpl.nasa.gov/progress_report/42-226/42-226F.pdf

Tanner, A., & Riley, A. (2003). Design and performance of a high-stability water vapor radiometer. Radio Science, 38(3). https://doi.org/10.1029/2002RS002673

Thornton, C., & Border, J. (2003). Radiometric tracking techniques for deep-space navigation. John Wiley & Sons.

Tortora, P., Crewe, S., Elgered, G., Graziani, A., Jarlemark, P., Loechner, U., et al. (2013). AWARDS: Advanced microwave radiometers for deep space stations. Space Communications. https://doi.org/10.3233/SC-130011

Tortora, P., Less, L., Bordi, J., Ekelund, J., & Roth, D. (2004). Precise Cassini navigation during solar conjunctions through multifrequency plasma calibrations. Journal of Guidance, Control, and Dynamics, 27(2), 251–257. https://doi.org/10.2514/1.997

Tortora, P., Zannoni, M., Hemingway, D., Nimmo, F., Jacobson, R. A., Less, L., & Parisi, M. (2016). Rhea gravity field and interior modeling from Cassini data analysis. Icarus, 264, 264–273. https://doi.org/10.1016/j.icarus.2015.09.022

Zannoni, M., Hemingway, D., Casajus, L. G., & Tortora, P. (2020). The gravity field and interior structure of Dione. Icarus, 345. 113713. https://doi.org/10.1016/j.icarus.2020.113713

Zannoni, M., & Tortora, P. (2013). Numerical error in interplanetary orbit determination software. Journal of Guidance, Control, and Dynamics, 36(4), 1008–1018. https://doi.org/10.2514/1.59294

Zumberge, J. F., Helfin, M. B., Jefferson, D. C., Watkins, M. M., & Webb, F. H. (1997). Precise point positioning for the efficient and robust analysis of GPS data from large networks. Journal of Geophysical Research: Solid Earth, 102(B3), 5005–5017. https://doi.org/10.1029/96JB03860