

PAPER

Data assimilation of GPS-ZTD into the RAMS model through 3D-Var: preliminary results at the regional scale

To cite this article: Alessandra Mascitelli *et al* 2019 *Meas. Sci. Technol.* **30** 055801

View the [article online](#) for updates and enhancements.

Data assimilation of GPS-ZTD into the RAMS model through 3D-Var: preliminary results at the regional scale

Alessandra Mascitelli^{1,2}, Stefano Federico¹, Marco Fortunato²,
Elenio Avolio³, Rosa Claudia Torcasio³, Eugenio Realini⁴,
Augusto Mazzoni², Claudio Transerici¹, Mattia Crespi²
and Stefano Dietrich¹

¹ ISAC-CNR, via del Fosso del Cavaliere 100, Rome, Italy

² Sapienza Università di Roma, via Eudossiana 18, Rome, Italy

³ ISAC-CNR, zona Industriale comparto 15, 88046 Lamezia Terme, Italy

⁴ Geomatics Research and Development s.r.l., Via Cavour 2, Lomazzo (CO), Italy

E-mail: alessandra.mascitelli@uniroma1.it

Received 2 October 2018, revised 25 February 2019

Accepted for publication 28 February 2019

Published 22 March 2019



Abstract

The knowledge of water vapour distribution is a key element in atmospheric modeling and considerable information, also at the local scale, can be derived from the GPS-ZTD (global positioning system-Zenith total delay) data.

This paper shows the assimilation of GPS-ZTD data into the RAMS@ISAC (Regional Atmospheric Modeling System at Institute of Atmospheric Sciences and Climate of the National Research Council) to improve the representation of the water vapour in the meteorological model.

The data assimilation system is based on 3D-Var (three-dimensional variational assimilation system) and it is applied to a network of 29 receivers located within the Lazio Region, Central Italy. All collected data are processed using the PPP (precise point positioning) method through RTKLIB, an open source program package for GNSS (Global Navigation Satellite Systems) Positioning. Among the GPS receivers, three are single frequency receivers, able to acquire $L1$ frequency only, so that it is necessary a preliminary reconstruction of $L2$ synthetic observations, which is achieved by a new original ground-based augmentation strategy. Results show remarkably that the single frequency receivers can be used the same way as geodetic receivers.

The RAMS@ISAC is run at 4 km horizontal resolution over central Italy and is nested, using one-way nesting, into a 10 km horizontal resolution run of the same model. The experiment was performed along to two months, from 28 July to 28 September 2017.

Results show that the GPS-ZTD data, assimilated by 3D-Var, have an important impact on the analysis of the water vapour field and the RMSE of ZTD and IWV (vertically integrated water vapour) is roughly halved for the analysis compared to the background.

The impact of the GPS-ZTD data assimilation is also evaluated for the very short term (VSF) forecast (1–3 h), obtaining an improvement of the ZTD and IWV RMSE for all three hours of forecast.

Keywords: geodesy, geomatics, meteorology, GPS, Zenith total delay, three-dimensional variational assimilation system, RAMS

(Some figures may appear in colour only in the online journal)

Introduction

Water vapour plays a crucial role in atmospheric processes that act over a wide range of temporal and spatial scales, from global climate to micrometeorology. Therefore, a good knowledge of water vapour distribution in space and time is a fundamental requirement for several atmospheric applications, especially weather forecasting.

Relatively cheap and easy information on water vapour distribution can be derived from Global Navigation Satellite Systems (GNSS). Usually, GNSS refer to series of available constellations of which GPS (global positioning system) is the most famous.

GNSS data can be very useful for the purpose of a multi-instrumental investigation; indeed, it was used to calibrate and validate other instruments (Campanelli *et al* 2018) or methodologies (Rohm *et al* 2011), but one of the most impactful application is related to the assimilation of GNSS data in numerical weather prediction models (for example: Vedel and Huang 2004, Faccani *et al* 2005, Poli *et al* 2007, Bennitt and Jupp 2012, Lindskog *et al* 2017).

Vedel and Huang (2004) assimilated GPS-ZTD into the HIRLAM (High Resolution Limited Area Model) using 3D-Var and found improvements for the forecast of geopotential height and high precipitation. Faccani *et al* (2005), using the MM5 model at 9 km horizontal resolution and 3D-Var to assimilate GPS-ZTD over Italy, found some improvements in the precipitation forecast during the transition from winter to spring.

Poli *et al* (2007) assimilated the GPS-ZTD by using 4D-Var and the Météo-France ARPEGE (Action de Recherche Petite Echelle Grande Echelle) global model. Results show the positive impact of the GPS-ZTD data assimilation on the forecast of synoptic-scale circulations and some improvement on the precipitation forecast in spring and summer. Following the experiment of Poli *et al* (2007), both Yan *et al* (2009) and Boniface *et al* (2009), both using a limited area model at about 2.5 km horizontal resolution, found positive impact of the assimilation of GPS-ZTD on precipitation forecast for heavy precipitation events occurred over France.

Bennitt and Jupp (2012) assimilated GPS-ZTD observations by both 3D-Var and 4D-Var using the Met Office NAE model (Rawlins *et al* 2007) at 12 km and 24 km horizontal resolutions, depending on the specific numerical experiment. The assimilation of GPS-ZTD increased the relative humidity at the levels considered for the analysis, resulting in an improvement of the cloud forecast. However, using the 4D-Var instead of 3D-Var had a limited impact on the forecast.

Lindskog *et al* (2017) performed GPS-ZTD data assimilation into the HARMONIE-AROME model run at 2.5 km horizontal resolution. The assimilation was performed by 3D-Var and improved the forecast up to one and half day, especially for humidity.

In this research, we show the assimilation of GPS-ZTD from both geodetic (dual-frequency) and single-frequency receivers in the Regional Atmospheric Modeling System model at the Institute for Atmospheric Sciences and Climate (RAMS@ISAC).

The data assimilation is performed by 3D-Var, extending the methodology of Federico (2013), and its impact on the analysis and very short-term forecast (VSF) of vertically integrated water vapour (IWV) is shown by a numerical experiment spanning two months, from 28 July to 28 September 2017.

The experiment is focused on the regional scale, over the Lazio Region in Central Italy, using a network of twenty-nine receivers. Three of them are single frequency receivers, i.e. acquire observation on $L1$ frequency only; the rationale behind their use is to improve the spatial resolution of the GPS-ZTD observations with reasonable budget. This aspect could be especially important in cases of morphologically complex terrain.

The paper is organized as follow: in section 2 we introduce the GPS observation and processing, the GPS network, the meteorological model, the 3D-Var method and the implemented observations quality control. Section 3 shows the results for the considered two months period and for the very short-term forecast. Conclusions are given in section 4, while details on the forward observation operator are given in appendix.

Data and methods

GPS observations and processing software

Water vapour content of the atmosphere, especially the lower layer (up to about 15 km) known as troposphere, affects GNSS signals by lowering their propagation velocities with respect to vacuum. A diminished speed results in a time delay of the signal propagation along the satellite-receiver path, that multiplied by the vacuum speed of light adds an extra-distance to the satellite-receiver geometrical one. This delay, called STD (slant total delay), is usually modeled through a parameter named ZTD (Zenith total delay), consisting of a hydrostatic (ZHD) and a wet (ZWD) part, multiplied by a mapping function (MF) accounting for the satellite elevation (Hofmann-Wellenhof *et al* 2007, Leick *et al* 2015).

If, from the positioning point of view, this delay is just a systematic error to be removed, it puts forward GNSS as a tool for the remote sensing of the tropospheric water vapour content.

To do this, different software and methods can be used; in this case we use the PPP (precise point positioning) technique, as implemented by RTKLIB, an open source program package for GNSS positioning (Takasu and Yasuda 2009). Another PPP solution, by the open source software goGPS (Herrera *et al* 2016), was employed to validate the correctness of our computations.

GPS network

The experiment is performed at the local scale over Lazio Region, in Central Italy, referring to the period 28 July–28 September 2017 and using only GPS observations. The

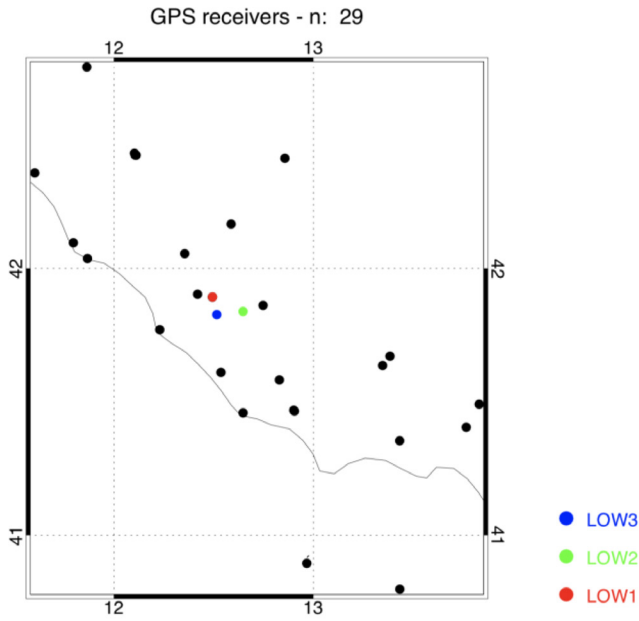


Figure 1. GPS permanent network.

network (figure 1) is composed by 26 geodetic receivers from three different permanent networks (ItalPos, Netgeo and Rete Lazio) and 3 single frequency receivers, which constitute a single frequency receivers pilot network; these receivers are located on the roofs of three buildings in Rome (figure 1), at inter-distances lower than 15 km:

1. Department of Civil, Constructional and Environmental Engineering of Faculty of Civil and Industrial Engineering, Sapienza University of Rome
2. Institute of Atmospheric Science and Climate, National Council of Research
3. National Institute of Geophysics and Volcanology

Ionospheric modelling for single frequency receivers: A new strategy

$L2$ observation for GNSS single frequency receiver is reconstructed through a new methodology, starting from the observations acquired by a unique geodetic class receiver (for which, of course, $L1$ and $L2$ are available) placed in its surroundings. The algorithm computes the time variation epoch-by-epoch of the ionospheric delay for each satellite in view from a geodetic class receiver through the geometry—free linear combination to its carrier phase observations ($L1$ e $L2$). The observation equation for the carrier phase for generic frequency (i), receiver (R) and satellite (S) can be written as follows.

$$L_i^S = \rho_R^S + c (dt_R - dt^S) - I_i^S + T_R^S + \lambda i N_i^S. \quad (1)$$

The terms represent, respectively, the geometric distance satellite/receiver, the difference between receiver and satellite clock offset, ionospheric and tropospheric delay and the multiplication of the wavelength and initial ambiguity term.

Omitting relative indices for the receiver and the satellite, the ionospheric delay $I1$ can be computed in the hypothesis no cycle slip occurred.

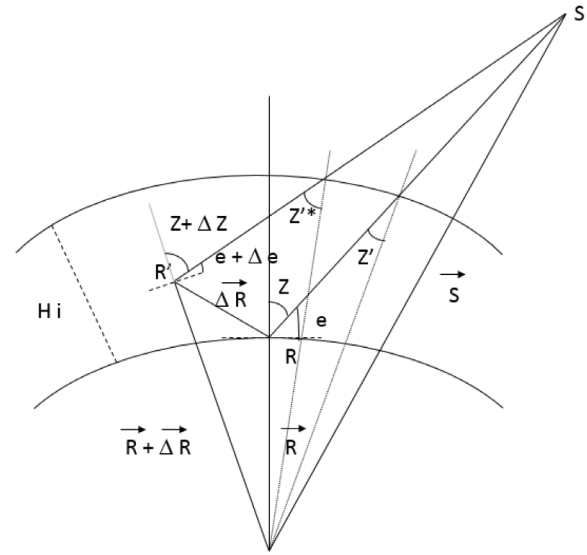


Figure 2. Schematic representation of the innovative methodology.

$$L1 - L2 = -I1 + I2 = -I1 + I1 \frac{f_1^2}{f_2^2}. \quad (2)$$

Where f_i represents the signal frequency. These ionospheric delays variations are mapped, using the concept of MF in Zenith direction ($Iz1$). Single Layer Models (SLMs) are commonly used in Geodesy to model Ionosphere by its thin shell approximation placed at a mean height of 350 km from the Earth surface. Furthermore, the use of a MF allows to project in Zenith direction ($Iz1$) the ionospheric delay $I1$ computed on the line of sight. In this methodology single-layer model (SLM) MF (Schaer 1999) is used; it assumes that the electron density field over a general GNSS station is spherically symmetric (figure 2). The relation between $Iz1$ and $I1$ follows.

$$Iz1 = I1 \cos Z'. \quad (3)$$

$Iz1$ derives from the combination of (2) and (3).

$$Iz1 = \frac{(L1 - L2)}{\frac{f_1^2 - f_2^2}{f_2^2}} \cos Z'. \quad (4)$$

Therefore, considering two consecutive epochs, the time variation of ionospheric content in Zenith direction $\delta Iz1$ can be obtained.

$$\delta Iz1 = \frac{\delta(L1 - L2)}{\frac{f_1^2 - f_2^2}{f_2^2}} \cos Z'. \quad (5)$$

In a similar way, it is possible to reconstruct the ionospheric delay $I1^*$ for the satellites in view from a generic GNSS single frequency receiver placed at a distance ΔR from the considered reference station (figure 2).

$$I1^* = \frac{Iz1}{\cos Z'^*} = \frac{(L1 - L2)}{\frac{f_1^2 - f_2^2}{f_2^2}} \frac{\cos Z'}{\cos Z'^*}. \quad (6)$$

In fact, using the MF allows to take in account the different path followed by the signal that, starting from the satellite S ,

is collected by two different receivers and, for this reason, it has a key role in the accuracy of the ionospheric model. At the end, the time variation of synthetic $L2^*$ observation, for each satellite in view from both the receivers at the same time, follows.

$$\delta L2^* = \delta L1^* - \delta(L1 - L2) \frac{\cos Z'}{\cos Z'^*}. \quad (7)$$

The term $\frac{\cos Z'}{\cos Z'^*}$ can be obtained starting from simple trigonometric considerations.

$$\cos Z' = \sqrt{1 - \left(\frac{|\vec{R}| \cos e}{|\vec{R}| + H_1} \right)^2} \quad (8)$$

$$\cos Z'^* = \sqrt{1 - \left(\frac{|\vec{R} + \vec{\Delta R}| \cos(e + \Delta e)}{|\vec{R}| + H_1} \right)^2}. \quad (9)$$

Where \vec{R} represents the Earth's radius, H_1 the ionospheric height above the terrestrial surface, e the elevation of the S respect to the GNSS dual frequency receiver R , $(e + \Delta e)$ the elevation of the satellite S respect to the GNSS single frequency receiver R' and $\vec{\Delta R}$ the distance between the two receivers. The same strategy can be applied in order to obtain the synthetic pseudorange transmitted on $L2$ frequency to perform a code-based positioning technique, such as single point positioning (SPP) or differential code-based positioning (DGNSS).

GPS processing

Starting from the dual frequency observational files (RINEX Version 3 format) collected by the geodetic receivers at 30 s rate, the PPP technique (Zumberge *et al* 1997), undifferenced phase observation processing, was applied using Ionospheric free combination in order to estimate both coordinates and ZTD values for each epoch, by daily processing sessions. Moreover, having used the *combined* post-processing option in RTKLIB (as implemented in RTKLIB (Takasu and Yasuda 2009)), it was not necessary to speed up the convergence time. As regards the ancillary products (ephemeris and clocks), we used precise products provided by the Center for Orbit Determination in Europe (CODE). For what concerns the software, we opted for RTKLIB software. An example of results is shown in figure 3.

The same procedure was followed for the three single frequency receivers, after the reconstruction of $L2$ synthetic observations by the new original ground-based augmentation strategy described before.

The validation procedure by other software was conducted only for single frequency devices, whereas for what concerns geodetic receivers, an explanatory test was conducted putting in relation ZTD data obtained by ROUN (one geodetic receiver belonging to the network used in this study) with official estimations provided by EUREF Permanent GNSS Network for another geodetic receiver, MOSE. These two

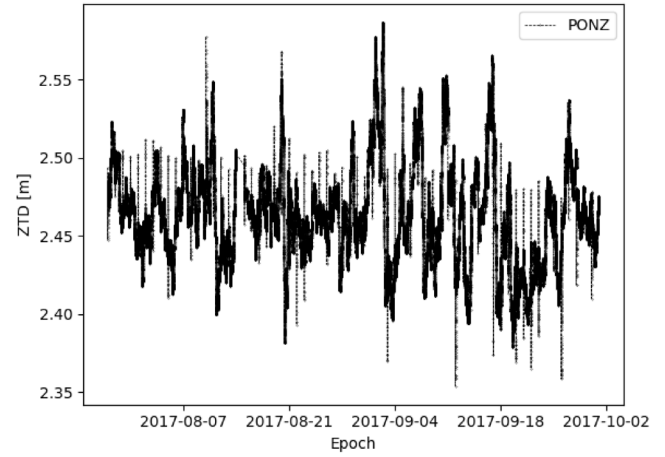


Figure 3. Example of ZTD output for PONZ geodetic receiver during the whole period.

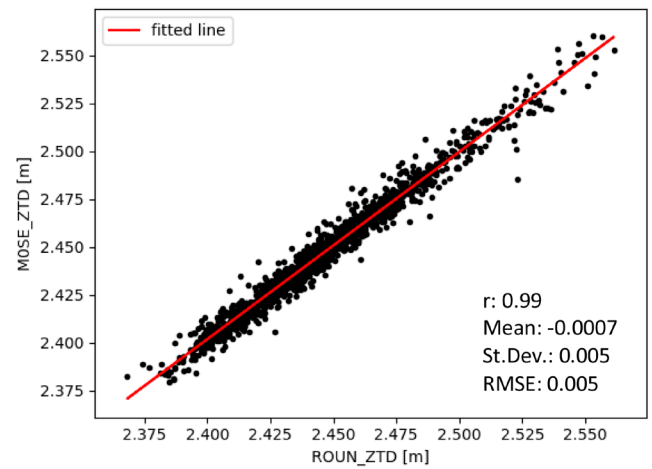


Figure 4. Comparison between ZTD estimation by ROUN and ZTD from EUREF official estimation of MOSE.

devices are co-located on the roof of Engineering Faculty of 'Sapienza' University of Rome and their comparison reveal great consistency, as shown in figure 4.

As regards the validation procedure by goGPS, it was conducted using a PPP batch processing at 30 s rate and ancillary products (ephemeris and clocks) by IGS (International GNSS Service); in the latter case, the reconstruction of $L2$ synthetic observations was obtained by SEID (satellite-specific epoch-differenced ionospheric delay) model (Zou *et al* 2010).

The ZTD estimated by RTKLIB extended Kalman filter (EKF) is modelled as a white noise, random walk parameter; goGPS applies the same model, but instead of processing observations sequentially as in an EKF, all observations are adjusted together in a least-squares system. To introduce the random walk noise, a first-order Tykhonov regularization is applied, similarly to the approach followed by the GNSS processing software Bernese (Dach *et al* 2015).

In this case study no outlier detection on ZTD estimates was employed.

The output comparison (figure 5), whose statistics are reported in table 1, shows the successful validation of computations. Therefore, given the consistency of the estimated

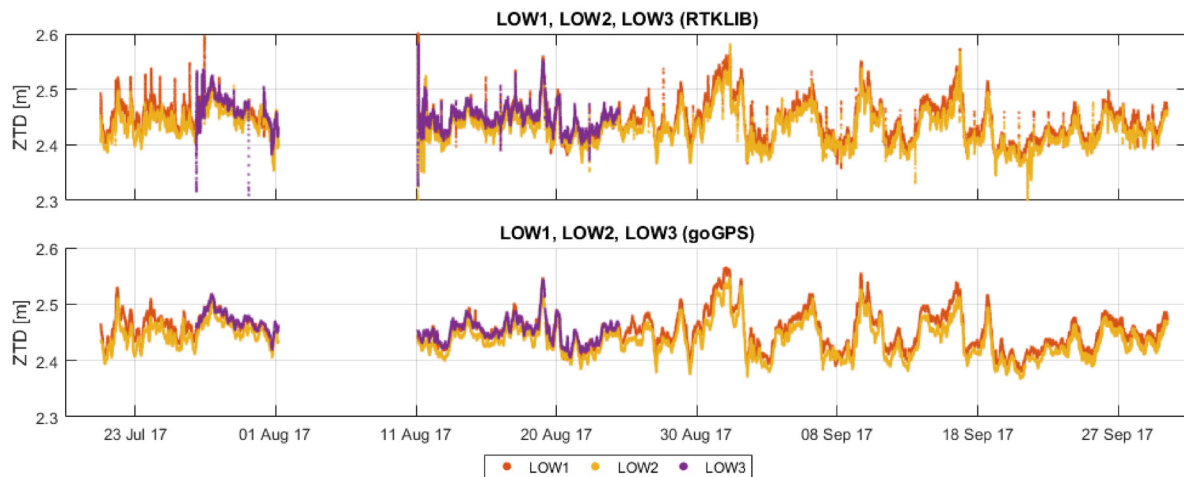


Figure 5. Comparison between single frequency receivers output during the whole period processed by RTKLIB (top) and goGPS (bottom).

Table 1. Statistics related to the difference between goGPS and RTKLIB results.

	LOW1	LOW2	LOW3
MEAN (mm)	1	0	-1
ST.DEV. (mm)	11	10	12

ZTDs at 1 centimeter level, it was chosen to assimilate the results obtained by RTKLIB; performing a correlation analysis on these latter with data obtained by geodetic receivers. As an example, the comparative analysis carried out between a single frequency receiver (LOW1) and a geodetic receiver (ROUN), both located on the roof of Engineering Faculty of ‘Sapienza’ University of Rome, is reported. As can be seen from the results (figure 6), LOW1 data show great consistency with those of ROUN, so as to show a greater agreement between them [RMSE = 0.63 cm], compared to the model [RMSE_LOW1 = 1.87 cm, RMSE_ROUN = 1.47 cm].

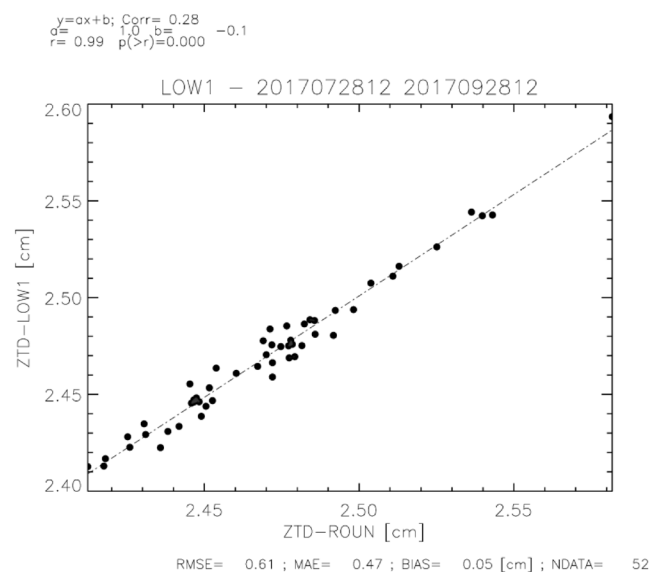


Figure 6. Comparison between ZTD from geodetic receiver (ROUN) and ZTD from single frequency receiver (LOW1).

RAMS model

We use RAMS model in a version maintained and developed, starting from the RAMS 6.0 model (Cotton *et al* 2003), at ISAC-CNR. This version implements the WRF Single-moment Microphysics Scheme 6 (WSM6, Hong and Pan 1996), whose details about the implementation in RAMS is described in Federico (2016), and predicts lightning (Federico *et al* 2014). Important for this work, it is the use of the 3D-Var analysis package described in the next section and developed at ISAC-CNR (Federico 2013). Table 2 resumes the full list of physical parameterization schemes used in this paper.

The grid configuration for the numerical experiments uses two one-way nested grids (table 2, figure 7). The first grid (referred as R10) has 10 km horizontal resolution and covers the Central Mediterranean area. The second grid (referred as R4) has 4 km horizontal resolution, extends over most of Italy, and it is centered over Central Italy. The vertical levels are 36 and extend from surface to the lower stratosphere while the vertical coordinate is terrain following.

Using the above configuration, we run the model at 10 km horizontal resolution starting at 00 UTC and performing a 15 h forecast on each day of the period considered. This forecast gives the initial and boundary conditions to the model at 4 km horizontal resolution (figure 8). More specifically at the 06 UTC on each day, the inner domain is initialized by the R10 model and performs a 9h forecast (R4 forecast). At 12 UTC, an analysis is made using as background the 6h short-term forecast of R4 and the ZTD data, estimated through GPS, collected at 12 UTC. Then a short-term 3h forecast is made, R4_ANL, and the vertically IWV and ZTD at 12, 13, 14 and 15 UTC are compared between R4 and R4_ANL, to quantify the benefits of the GPS-ZTD data assimilation in the analysis/forecast. We do not consider longer time scales because the innovations introduced by the GPS-ZTD data are rapidly advected out of the domain where the GPS receivers are located.

Table 2. List of physical parameterizations used for RAMS in this paper.

Physical parameterization	Selected scheme
Parametrized cumulus convection	Modified Kuo scheme to account for updraft and downdraft (Molinari and Corsetti 1985). The scheme is applied to R10 only
Explicit precipitation parameterization	Bulk microphysics with six hydrometeors (cloud, rain, graupel, snow, ice, water vapour). Describer in Hong and Lim (2006)
Exchange between the surface, the biosphere and atmosphere.	LEAF3 (Walko et al 2000). LEAF includes prognostic equations for soil temperature and moisture for multiple layers, vegetation temperature and surface water, including dew and intercepted rainfall, snow cover mass and thermal energy for multiple layers, and temperature and water vapour mixing ratio of canopy air
Sub-grid mixing	The turbulent mixing in the horizontal directions is parameterized following Smagorinsky (1963), vertical diffusion is parameterized according to the Mellor and Yamada (1982) scheme, which employs a prognostic turbulent kinetic energy
Radiation scheme	Chen and Cotton (1983). The scheme accounts for condensate in the atmosphere, but not for specific optical properties of ice hydrometeors

3D-Var data assimilation

We extended the 3D-Var analysis scheme of the RAMS model (Federico 2013) to assimilate the GPS-ZTD. The aim of the 3D-Var is to minimize the cost-function $J(x)$ given by:

$$J(x) = \frac{1}{2}(x - x_b)^T B^{-1}(x - x_b) + \frac{1}{2}(y_o - H(x))^T R^{-1}(y_o - H(x)). \quad (10)$$

Where x_b is the background state vector, B is the background error matrix, y_o is the observation vector, R is the observation error matrix and H is the forward observation operator, transforming the state vector into the observed variable (simulated ZTD in correspondence of the GPS receivers in this case).

x_b is a column vector whose length is $nvar \times nxyz$, where $nxyz$ is the number of the RAMS grid points, i.e. $nxyz = npx \times nny \times nnz$, and $nvar$ is 2, the mixing ratio of water vapour and temperature. Stated in other terms we order the model values of water vapour mixing ratio and temperature in a column vector, where the first $nxyz$ components refer to water vapour mixing ratio and the second $nxyz$ values refer to temperature. Each of the variables is ordered starting from the SW corner of the first model vertical level and moving first towards east and then to the north. Once the first level is completed we pass to the second level, and so on until the NE grid corner of the last level is reached.

The B matrix contains a description of the model error as a function of the position and level and is decomposed in the three spatial directions as detailed below in this section.

y_o is a column vector whose length is np , i.e. the number of observations available at the assimilation time after the quality control procedure has been applied to the ZTD data, estimated through GPS, from here on out considered as observations (see next Section for the details on the quality control applied).

The H operator, transforming the state vector into ZTD, is described in appendix.

In the RAMS-3D-Var, the cost-function is implemented in the incremental form (Courtier et al 1994) by introducing the transform $x' = Uv$, where $x' = x - x_b$ and U is chosen to satisfy the relationship $B = UU^T$. Using this transform, the cost-function (1) can be written as follows:

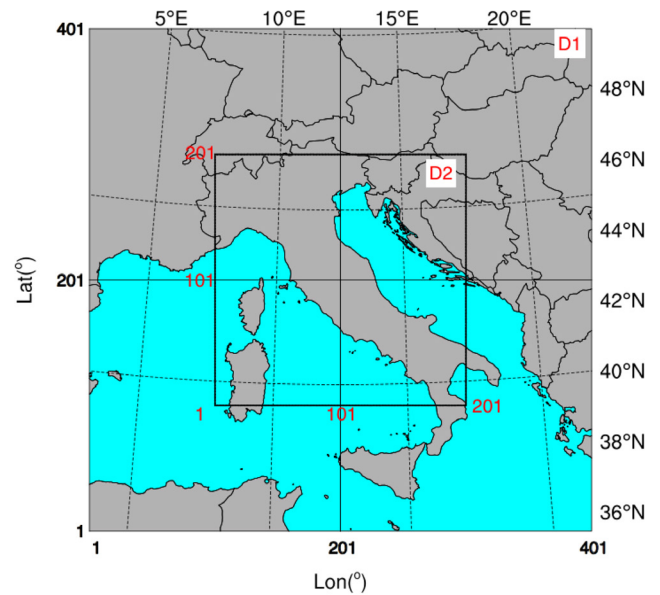


Figure 7. Domains of R10 (D1), R4 and R4_ANL (D2) forecasts. The most important parameters of the domains are given in table 2.

$$J(v) = \frac{1}{2}v^T v + \frac{1}{2}(y_o - HUv)^T R^{-1}(y_o - HUv) \quad (11)$$

where H is the linearization of the forward operator H and $y'_o = y_o - H(x)$.

In general, the RAMS 3D-Var is able to assimilate vertical profiles of temperature, relative humidity and winds, as discussed in Federico (2013). New features of the 3D-Var include the assimilation of radar reflectivity and GPS-ZTD, while lightnings are assimilated through nudging. Finally, surface observations are assimilated by Optimal Interpolation (Kalnay 2003, Federico 2011). In this paper, however, only GPS-ZTD estimated through GPS are considered observations for the RAMS 3D-Var and are assimilated.

To discuss more in detail the transformation $x' = Uv$ we note that the B matrix is decomposed in the three spatial directions $B = B_z B_y B_x$. The B_x and B_y matrices depend on the spatial distance between two grid-points and represent the

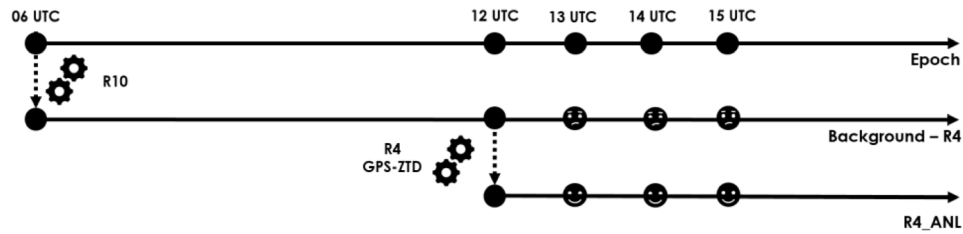


Figure 8. Temporal scheme of the simulations. At 6 UTC of each day a 9 h forecast is performed (R4), using the R10 forecast as initial and boundary conditions. At 12 UTC an analysis is made and a short-term forecast (3 h) starting from this analysis is performed (R4_ANL).

background error correlation length scale. For these matrices, a Gaussian form is assumed:

$$B_x = e^{-\frac{x^2}{2L_x^2}}; B_y = e^{-\frac{y^2}{2L_y^2}} \quad (12)$$

where L_x and L_y are the background error length-scales that are determined by the National Meteorological Center (NMC) method (Parrish and Derber 1992). The determination of the length-scales was performed considering two sets of short term forecasts, 12 h and 24 h respectively, verifying at 12 UTC on each day for the whole period and applying to these forecasts the NMC method (Parrish and Derber (1992), see Barker *et al* (2004) and Federico (2013), for the details on the methodology). The results gave length-scales ranging between 20 and 30 km, depending on the model level.

The B_x and B_y matrices are symmetric and positive defined and can be decomposed in the eigenvectors and eigenvalues matrices. For example, $B_x = VL V^T$, where L is the matrix of eigenvalues (i.e. a square diagonal matrix having along the diagonal the eigenvalues and zeroes elsewhere) and V is the matrix whose columns are the eigenvectors. Using this decomposition, the transform U_x is: $U_x = VL^{1/2}$, and similarly for U_y .

The B_z matrix is also computed by the NMC method. In addition to the correlation length in the vertical direction, B_z takes into account the model error at each level. For the water vapour mixing ratio, this error is about 2 g kg^{-1} at the surface and decreases with height. In particular, it is larger than 0.5 g kg^{-1} below 4 km, and less than 0.2 g kg^{-1} above 5 km.

For temperature, the model error is more than 2 K at the surface, and then it decreases with height being less than 0.5 K above 2 km. The temperature error is less than 0.5 K up to 11 km, where it increases reaching 1.5 K at 13 km, showing the impact of the correct location of the tropopause on the model error. Above 14 km the temperature error decreases with height up to the model top.

It is noted that cross correlations among variables are not considered in this paper and the analysis of GPS-ZTD data impacts the water vapour mixing ratio and the temperature of the model, leaving unchanged other variables. Moreover, the impact of the data assimilation on temperature is very small (less than 0.05 K).

The errors of the GPS receivers are assumed uncorrelated and the R matrix is diagonal. The error of each GPS-receiver is 1 cm and, when more than one observation is found inside a grid box, only the first observation is considered. Among the twenty-nine receivers considered in this paper four were

excluded because they are in the same grid box of other receivers (two in Viterbo, VIRB and VITE, one in Latina, LTNA, and the single frequency receiver LOW1).

The minimization of the cost-function is performed by the conjugate-gradient method (Press *et al* 1992). Finally, details on the H operator are given in appendix.

Bias correction and data quality control

In a data assimilation system, the observations are assumed unbiased (Lorenc 1986). To ensure this requirement, we computed the bias between the background and the observations over the whole period and determined its value for each receiver. The bias was then removed from the observations.

After the bias correction, we introduced two quality controls. Our experiment is at 4 km horizontal resolution and the representation of local orography can be rather different from the reality for specific receivers located in complex terrain. To avoid large extrapolation/interpolation of the model output to the receiver location we excluded all the receivers having a height that differs by more than 300 m compared to the model surface height. This left out 4 receivers belonging to the used permanent network over the Lazio Region, leaving the 29 receivers of the network shown in figure 1. As stated in the previous section, among the 29 receivers four were discarded because occurring in the same grid box of other receivers.

To avoid introducing excessive departure between the observations and the background, an observation was discarded if its difference with the background was larger than 4 cm. This value, which is more conservative than that used in other studies (Bennett and Jupp 2012), was found by trials and errors, and ensured a reasonable performance of the analysis for days when the background performance was particularly poor.

Results

Analysis

In this section, we quantify the impact of the analysis at each station for the whole period. As stated in the previous section, when more than one receiver fall in the same grid cell only one of them is assimilated, however the errors are quantified for all receivers both assimilated and non-assimilated.

In doing the analyses the single frequency receivers were considered the same way as the geodetic receivers. There are two main reasons for this choice. First, when considering

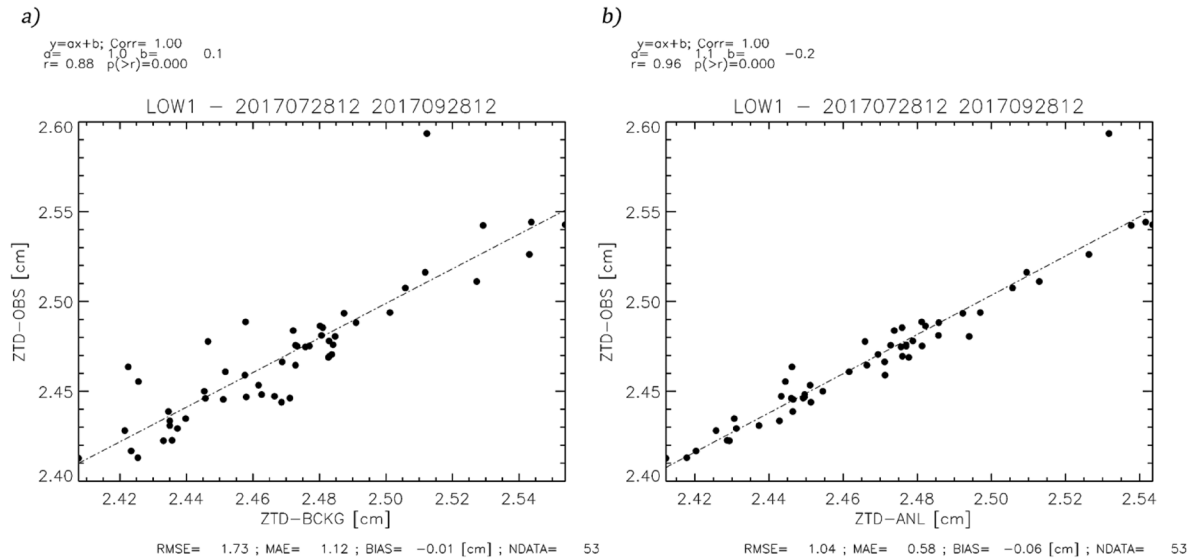


Figure 9. Comparison between the background (figure (a)) and analyses (figure (b)) with the observations of the LOW1 receiver. For this experiment none of the single frequency receivers were assimilated.

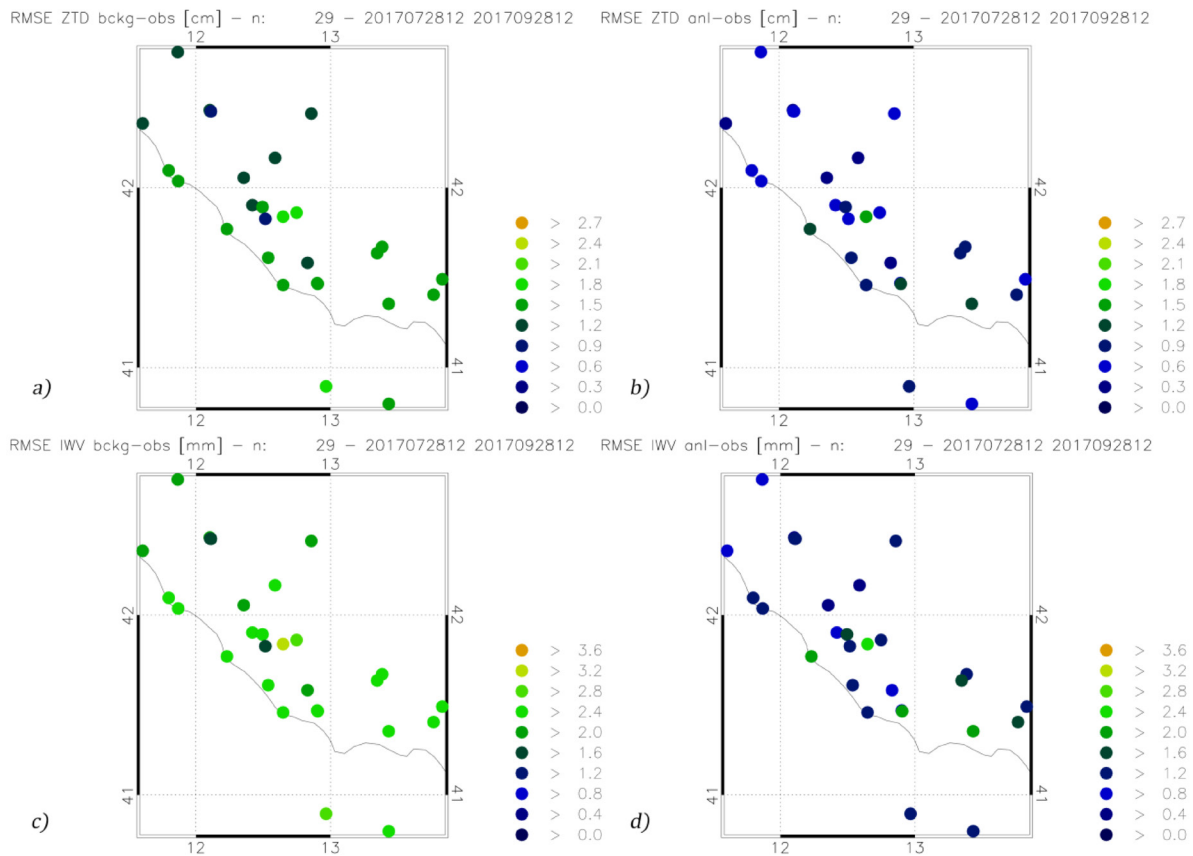


Figure 10. (a) RMSE of the ZTD for the background (R4) computed over the whole period; (b) as in (a) for the analysis (R4_ANL); (c) RMSE of the IWV for the background (R4) computed over the whole period; (d) as in (c) for the analysis (R4_ANL).

the performance of the background against observations, the single frequency receivers did not show any different or specific behavior compared to other receivers (see figures 8(a) and (b)). Second, an initial test was done in which the analyses were performed considering only the geodetic receivers, i.e. without assimilating the single frequency receivers. This test showed that the single frequency receiver observations were

closer to the analysis than to the background confirming, indirectly, the possibility to use them the same way as geodetic observations in performing the analysis. This is shown for the LOW1 receiver and for the ZTD in figure 9. The RMSE of the background is 1.74 cm, which reduces to 1.06 cm for the analysis. The correlation coefficient is 0.88 for the background and 0.96 for the analyses.

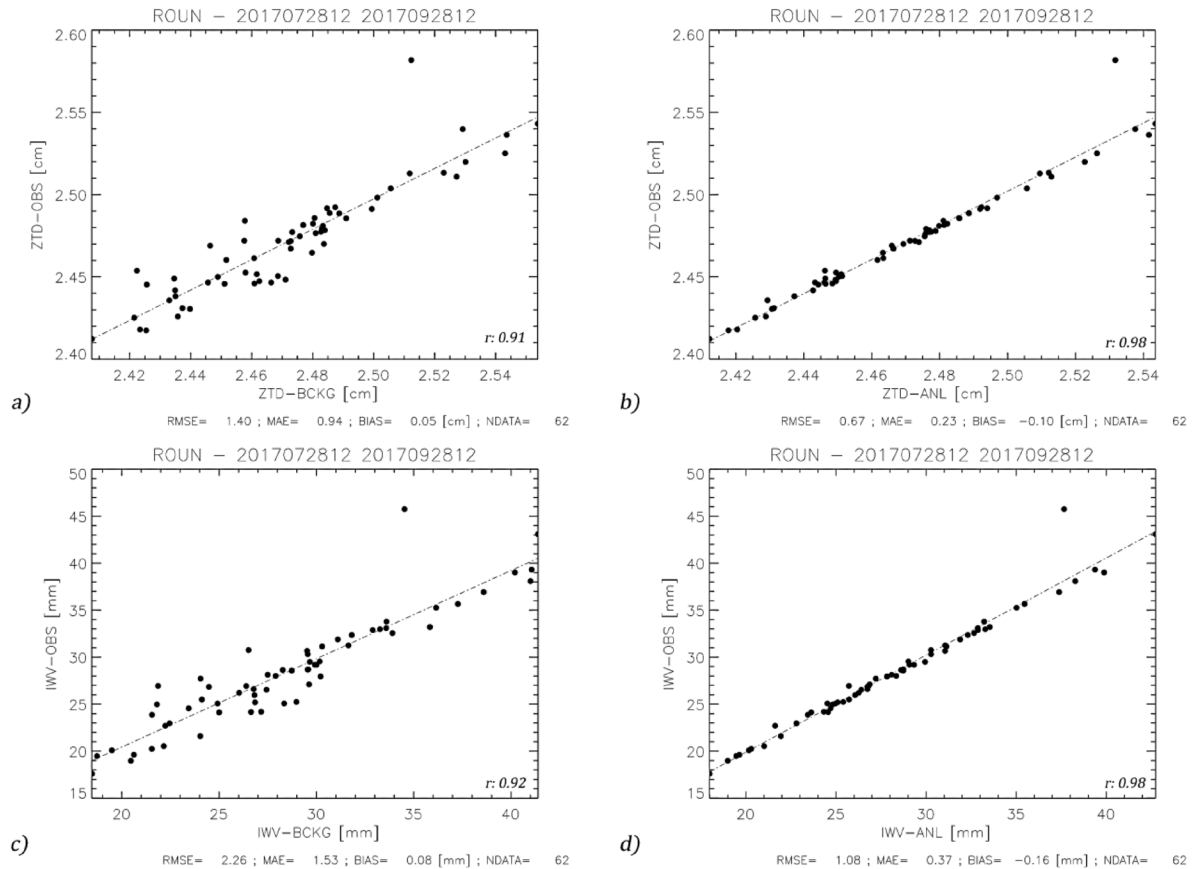


Figure 11. (a) Correlation between the background and the observations for ROUN for ZTD; (b) as in (a) for the analysis; (c) correlation between the background and the observations for ROUN for IWV; (d) as in (c) for the analysis.

Figure 10(a) shows the RMSE of the GPS-ZTD computed for the background and for the whole period at 12 UTC. Its value ranges between 0.9 cm and 2.1 cm, depending on the station. After the analysis, the error is considerably reduced, with values ranging from 0.3 cm to 1.8 cm. The ZTD error is reduced for all stations. Similar considerations apply for the Bias (not shown). In particular, the absolute value of the bias decreases for most stations (twenty-one out of twenty-two) for R4_ANL compared to R4.

The correlation coefficient between R4_ANL and the observations (not shown) increases for receivers compared to that between the background and the observations and all the correlation coefficients between R4_ANL and the observations are larger than 0.9.

The performance of the analysis is quantified also for the IWV, which is computed from the ZTD following Bevis *et al* (1992). The error for the background is shown in figure 10(c), which shows errors ranging from 2.0 mm to 3.6 mm, depending on the station. The analysis modifies the IWV showing an important impact (figure 10(d)), with errors ranging from 0.1 mm to 2.4 mm. For the bias and the correlation coefficient (not shown) similar considerations to the ZTD apply.

It is interesting to examine the behavior of the analysis for a specific site for the whole period. We choose the ROUN receiver (the position of this station is very close to the LOW1 receiver in figure 1 and the black dot is not visible because it

is under the red-dot of the LOW1 receiver) because it is representative of the results obtained for most receivers.

Figure 11(a) shows the correlation, for the whole period at 12 UTC, of the ZTD for the background (x -axis) and observations (y -axis). The correlation coefficient is 0.91, demonstrating the ability of the background to follow the evolution of the ZTD from day to day. The RMSE is 1.4 cm (figure 8(a)). There are also outliers of the ZTD estimated by the background, which occur for specific days when the background performance was particularly poor. The correction given by the analysis is significant, as shown in figure 11(b). The correlation coefficient is 0.98 and the RMSE is 0.67 cm, almost halved compared to the background value.

Figures 11(c) and (d) show the impact of the analysis on the IWV. The RMSE for the background is 2.3 mm and the correlation coefficient is 0.92. For the analysis the RMSE is 1.1 mm and the correlation coefficient 0.98.

In conclusion all the above results show the important and positive impact of the ZTD data assimilation to define the analysis fields that can be used to initialize the model.

Impact on the very short-term forecast

In this section we show the impact of the assimilation of the GPS-ZTD on the time-series of both the analysis and very short-term forecast (VSF, 1–3 h) of the ZTD and IWV

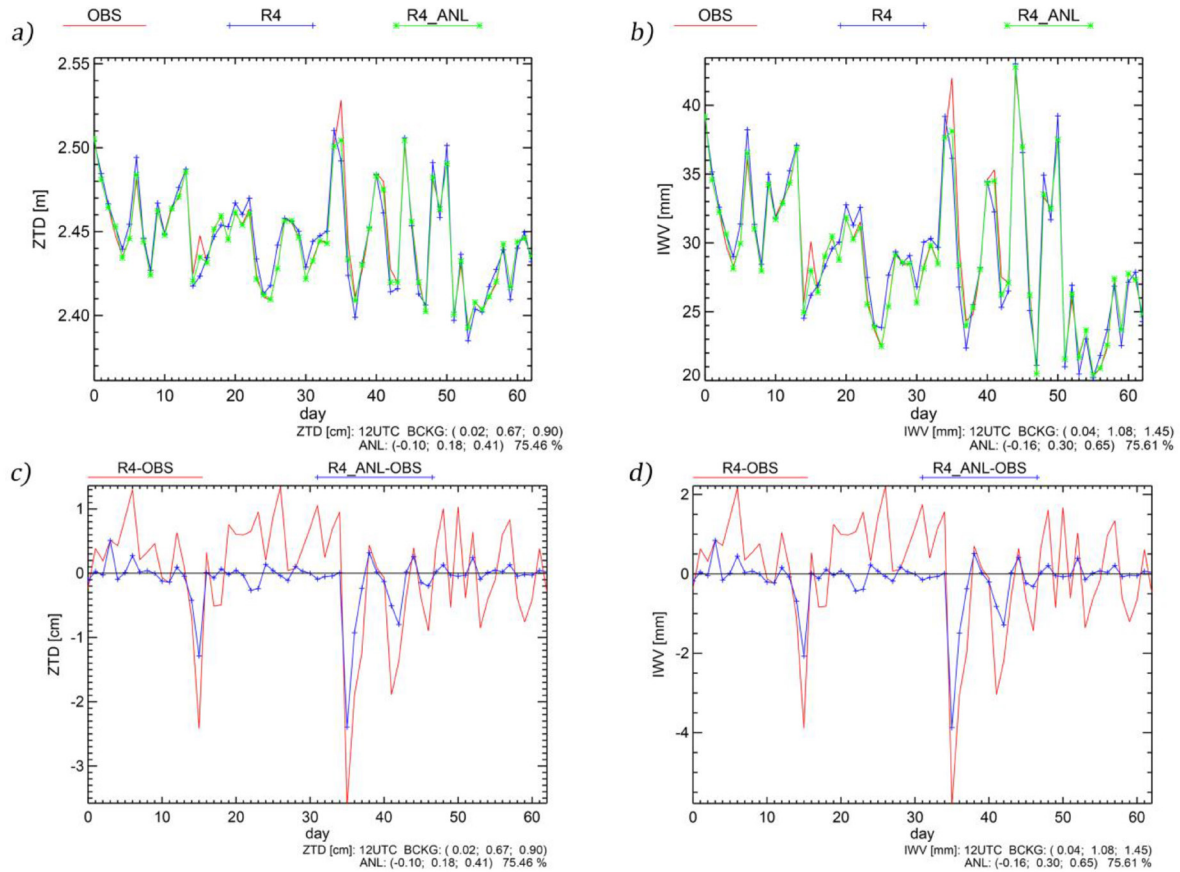


Figure 12. Time series of the R4, R4_ANL and observations (OBS) for ZTD (figure (a)), IWV (figure (b)) and for the differences between R4 and observations and between R4_ANL and observations for ZTD (figure (c)) and IWV (figure (d)).

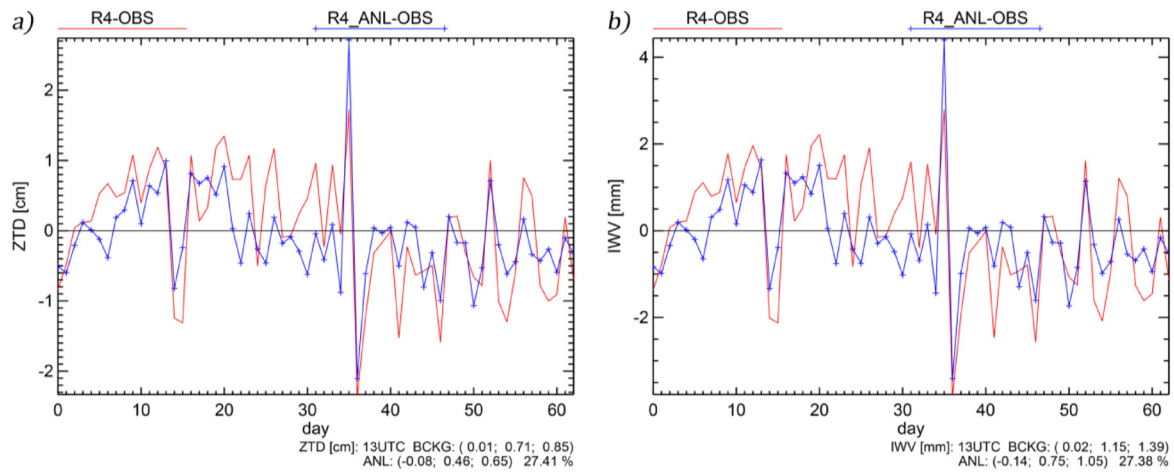


Figure 13. Time series of the differences between R4 and observations and between R4_ANL and observations for ZTD (figure (a)) and IWV (figure (b)). Verification at 1 h forecast.

averaged over all the receivers. For the RAMS simulations, i.e. R4 and R4_ANL, the fields of ZTD and IWV are first interpolated to the receivers' positions, and then the average over all receivers is considered. For the observations we consider the average over all the receivers.

Figure 9(a) shows the results for the ZTD at the analysis time, i.e. 12 UTC. There is a clear day-to-day dependence,

which is caused by the variability of the water vapour over the study area. For some days (for example day 15 and 35), when the error of the control simulation (R4) is larger, the impact of the GPS-ZTD data assimilation is apparent, however for most days it is difficult to see the difference between the simulations and observations. A similar consideration applies for the IWV (figure 9(b)). The time series of IWV for R4 and R4_ANL

Table 3. Basic parameters of the RAMS (R10, R4) grids. NNXP is the number of grid points in the WE direction, NNYP is the number of grid-points in the SN direction, NNZP is the number of vertical levels, DX is the size of the grid spacing in the WE direction, DY is the grid-spacing in the SN direction. Lx, Ly, and Lz are the domain extensions in the NS, WE, and vertical directions. CENTLON and CENTLAT are the coordinates of the grid centres. D1 refers to the first RAMS domain, D2 to the second domain.

	D1, R10	D2, R4/R4_ANL
NNXP	301	201
NNYP	301	201
NNZP	36	36
Lx	3000 km	800 km
Ly	3000 km	800 km
Lz	~22400 m	~22400 m
DX	10 km	4 km
DY	10 km	4 km
CENTLAT (°)	43.0 N	43.0 N
CENTLON (°)	12.5 E	12.5 E

Table 4. Bias, MAE and RMSE for the analysis and for the three hours of forecast. Statistics are computed for the whole period and are shown for ZTD (second and third column) and for IWV (fourth and fifth columns).

Time	R4, ZTD (cm)			R4_ANL, ZTD (cm)			R4, IWV (mm)			R4_ANL, IWV (mm)		
	Bias	MAE	RMSE	Bias	MAE	RMSE	Bias	MAE	RMSE	Bias	MAE	RMSE
12 UTC	0.02	0.67	0.90	−0.10	0.18	0.41	0.04	1.08	1.45	−0.16	0.30	0.65
13 UTC	0.01	0.71	0.85	−0.08	0.46	0.65	0.02	1.15	1.39	−0.14	0.75	1.05
14 UTC	0.20	0.76	0.96	0.14	0.54	0.80	0.34	1.25	1.56	0.23	0.88	1.30
15 UTC	0.12	0.75	0.90	0.07	0.56	0.71	0.21	1.23	1.47	0.12	0.91	1.16

shows differences between the two simulations of the order of 1 mm (3%–5% of the IWV depending on the day) and for some days larger than 2 mm (6%–10% of IWV depending on the day). This variation is not negligible because it represents an average over all the receivers and the differences are higher for specific locations.

The difference between the analysis and background for ZTD and IWV before and after the analysis can be represented more clearly considering the differences between R4 and GPS-derived ZTD and between R4_ANL and GPS-derived ZTD (figure 12(c) for ZTD and figure 12(d) for IWV). This representation highlights the improvement of the analysis compared to the background, because the blue line is closer to zero than the red line. The RMSE for ZTD is 0.90 cm for R4 and 0.41 for R4_ANL. For the IWV the RMSE after the analysis is 0.65 mm, while it is 1.45 mm for the background.

The differences between R4 and observations and R4_ANL and observations at 13 UTC, i.e. for the 1 h forecast, show an improvement of the forecast starting from the GPS-ZTD analysis for both ZTD (figure 13(a)) and IWV (figure 13(b)). At this time, there are some days when the forecast starting from the analysis is worse than the background, nevertheless the forecast at 1 h is improved for most days. This is confirmed by the statistics taken over the whole period at 13 UTC (table 3). For the ZTD, the RMSE is 0.85 cm for R4 and 0.65 cm for R4_ANL. For IWV, the RMSE at 13 UTC for R4 is 1.4 mm, while it is 1.1 mm for R4_ANL.

The forecasts at 14 and 15 UTC show characteristics similar to those of 13 UTC, and the forecast starting from the 12 UTC analyses has a lower error compared to the background. The statistics are detailed in table 4.

The RMSE is reduced by more than 10% of its value for both ZTD and IWV, showing a non-negligible impact of the GPS-ZTD data assimilation on the VSF.

Conclusions

This paper presents an extension of the RAMS 3D-Var data assimilation system to assimilate GPS-ZTD. The method is applied to the period 28 July–28 September 2017 using 29 GPS-receivers over the Lazio Region, Italy. GPS-ZTD is derived using the Precision Point Positioning technique implemented in RTKLIB.

The GPS network is composed by 26 geodetic receivers and by 3 single frequency receivers. The aim to extend the geodetic network with single frequency receivers is to cheaply enhance the representation of the atmospheric IWV, which is an important issue over complex terrain, as that considered in this study. Moreover, the assimilation of GPS-ZTD at the local scale could be important to better predict the precipitation at the local scale.

A preliminary study with single frequency receivers shows that they can be used in the data assimilation system because: (a) their error statistics did not show any specific behavior compared to other receivers; (b) the comparison between ZTD of the single frequency receivers with both the background and the analyses realized using only geodetic receivers showed a better agreement with the analyses.

Results show the important impact of the analyses on the ZTD and IWV fields. In particular, the RMSE of the analysis is more than halved compared to the background for most stations and the correlation coefficient between the analysis and

the observations is larger than 0.9 for most stations. Similar results were found for IWV.

Because the area covered by the GPS receivers is rather small, the impact of the 3D-Var data assimilation on the forecast was limited to the very short range (1–3 h). Results shows that the assimilation of GPS-ZTD reduced the RMSE error, computed over the whole period, by more than 10% of the background RMSE for both ZTD and IWV.

While the above results are encouraging, they are preliminary, and there are issues that need to be investigated in the future. Among them the following are highlighted:

- A) The impact of the GPS-ZTD data assimilation must be studied for other parameters than IWV, in particular for precipitation. A numerical study over the whole Italy will be performed in the near future to study the impact of the GPS-ZTD data assimilation on the precipitation forecast for the HyMeX-SOP1 (HYdrological cycle in the Mediterranean Experiment-Special Observing Period 1), held between 5 September and 6 November 2012. For this period a database of GPS-ZTD observation of a dense network of receivers is available (Bock *et al* 2016), as well as a detailed precipitation database. Both datasets are available through the HyMeX database (<http://mistrals.sedoo.fr>).
- B) Considering that the average distance between two GPS receivers of the dense network used in this work is about 20 km, an explicit treatment of the observation spatial error correlation is needed to exploit the increased density of GPS network with single frequency receivers and to take into account for non-homogeneous density of GPS receivers. Considering the applicability of receivers not homogeneously distributed into modelling data assimilation procedures, (Lindskog *et al* 2017) (as well as Sánchez-Arriola *et al* 2016), in testing the data assimilation of GNSS ZTDs by the Nordic GNSS Analysis Centre (NGAA), found that potential improvements can be found reducing thinning distance of the ZTD observations and, therefore, using more data. At the same time, they also found an increased forecast bias for humidity, likely caused by the increased influence of correlation errors. The horizontal observation error correlations will be considered in future works to optimise the use of non-homogeneous high-resolution GPS-ZTD data. Future studies will be conducted in this direction using, for example, the methodology of Bormann and Bauer (2010).

Acknowledgments

Part of the computations of this paper were performed on the ECMWF computing facilities through the project SPITFEDE. Aeronautica Militare and ECMWF are acknowledged for the access to MARS (Meteorological Archive and Retrieval System). The NetGEO GNSS Network is acknowledged for having provided the rinex data free of charge. This work was partially funded by the agreement between CNR-ISAC and the Italian Department of Civil Protection.

Appendix

The forward observation operator H

The forward observation operator computes the ZTD from the model variables. The ZTD can be expressed as:

$$H = \text{ZTD} = 10^{-6} \int_{z=0}^{z=\infty} N dz. \quad (\text{A.1})$$

Where z is the model height above the surface and N is the scaled-refractivity. N can be expressed as:

$$N = \frac{k_1 p_d}{T} + \frac{k_2 p_v}{T} + \frac{k_3 p_v}{T^2}. \quad (\text{A.2})$$

Where p_d is the partial pressure of dry-air, p_v is the water vapour partial pressure, T is the temperature and k_1 , k_2 , and k_3 are refractivity constants (Bevis *et al* 1992).

Substituting $p_d = p - p_v$ into equation (A.2), where p is the total pressure, and combining the second and third term of equation (A.2) (Bevis *et al* 1992), the scaled-refractivity can be expressed as:

$$N = \frac{ap}{T} + \frac{bp_v}{T^2} \quad (\text{A.3})$$

where the constants a and b are, respectively, 77.6 Kh Pa^{-1} and $3.73 \times 10^5 \text{ K}^2 \text{h Pa}^{-1}$. The expression (A.3) is used in our implementation of GPS-ZTD data assimilation.

RAMS uses an Arakawa-C grid in the vertical, where the thermodynamic and moisture variables are computed at the centre of the grid-box. In this configuration, all the terms in the RHS of equation (A.3) are defined at the half-levels and the computation of the integral of equation (A.1) is straightforward. In particular, the contribution of the i th model layer, whose vertical heights are between z_i and z_{i+1} , to the ZTD is given by:

$$\Delta \text{ZTD}_j = 10^{-6} N_j (z_{j+1} - z_j) \quad (\text{A.4})$$

and

$$N_j = \frac{ap_j}{T_j} + \frac{bp_{vj}}{T_j^2} \quad (\text{A.5})$$

and pressure is computed from the Exner function, which is a dependent variable of RAMS. Equations (A.4) and (A.5) are computed from the model surface to the model top and the contribution of each level is summed to give the ZTD of the modelled layer of atmosphere, i.e. from surface to about 22 km (table 2). Similarly to Bennitt and Jupp (2012), we account for the contribution to the ZTD from the model top to the top of the atmosphere.

This ZTD above the model top can be accounted by neglecting the second term in equation (A.5) above the model top, i.e. assuming that the contribution of water vapour to the ZTD is negligible above the model top, and assuming the hydrostatic equilibrium. Integrating equation (A.1) from the model top to the top of the atmosphere we have:

$$\text{ZTD}_{\text{TOP}} = \int_{z_{\text{TOP}}}^{\infty} N dz = 10^{-6} \int_0^{P_{\text{TOP}}} \frac{ap}{T} \frac{dp}{\rho g} = \frac{10^{-6} a R}{g} P_{\text{TOP}}. \quad (\text{A.6})$$

Where ρ is the air density, p_{TOP} is the pressure at the model top, g is the gravity acceleration, R is the gas-constant for dry air and we used the equation of state of the dry air.

Other authors, for example Poli *et al* (2007), account for the (small) contribution of ZTD_{TOP} in the bias correction.

ORCID iDs

Alessandra Mascitelli  <https://orcid.org/0000-0003-2925-0006>

Augusto Mazzoni  <https://orcid.org/0000-0002-1059-6675>

Stefano Dietrich  <https://orcid.org/0000-0003-3808-365X>

References

- Barker D M, Huang W, Guo Y-R and Xiao Q N 2004 A three-dimensional variational data assimilation system for MM5: implementation and initial results *Mon. Weather Rev.* **132** 897–914
- Bennitt G V and Jupp A 2012 Operational assimilation of GPS Zenith total delay observations into the Met office numerical weather prediction models *Mon. Weather Rev.* **140** 2706–19
- Bevis M, Businger S, Herring T A, Rocken C, Anthes R A and Ware R H 1992 GPS meteorology: remote sensing of atmospheric water vapor using the global positioning system *J. Geophys. Res.* **97** 15787–801
- Bock O, Bosser P, Pacione R, Nuret M, Fourrié N and Parracho A 2016 A high-quality reprocessed ground-based GPS dataset for atmospheric process studies, radiosonde and model evaluation, and reanalysis of HyMeX special observing period *Q. J. R. Meteorol. Soc.* **142** 56–71
- Boniface K *et al* 2009 Impact of high-resolution data assimilation of GPS Zenith delay on Mediterranean heavy rainfall forecasting *Ann. Geophys.* **27** 2739–53
- Bormann N and Bauer P 2010 Estimates of spatial and inter-channel observation error characteristics for current sounder radiances for NWP, part I: methods and application to ATOVS data *Q. J. R. Meteorol. Soc.* **136** 1036–50
- Campanelli M *et al* 2018 Precipitable water vapour content from ESR/SKYNET sun–sky radiometers: validation against GNSS/GPS and AERONET over three different sites in Europe *Atmos. Meas. Tech.* **11** 81–94
- Chen C and Cotton W R 1983 A one-dimensional simulation of the stratocumulus-capped mixed layer *Bound. Layer Meteorol.* **25** 289–321
- Cotton W R *et al* 2003 RAMS 2001: current status and future directions *Meteorol. Atmos. Phys.* **82** 5–29
- Courtier P, Thépaut J N and Hollingsworth A 1994 A strategy for operational implementation of 4D-Var, using an incremental approach *Q. J. R. Meteorol. Soc.* **120** 1367–87
- Dach R, Lutz S, Walser P and Fridez P 2015 *Bernese GNSS Software Version 5.2* (Bern: Astronomical Institute, University of Bern)
- Faccani C, Ferretti R, Pacione R, Paolucci T, Vespe F and Cucurull L 2005 Impact of a high density GPS network on the operational forecast *Adv. Geosci.* **2** 73–9
- Federico S 2011 Verification of surface minimum, mean, and maximum temperature forecasts in Calabria for summer 2008 *Nat. Hazards Earth Syst. Sci.* **11** 487–500
- Federico S 2013 Implementation of a 3D-Var system for atmospheric profiling data assimilation into the RAMS model: initial results *Atmos. Meas. Tech.* **6** 3563–76
- Federico S 2016 Implementation of the WSM5 and WSM6 single moment microphysics scheme into the RAMS model: verification for the HyMeX-SOP1 *Adv. Meteorol.* **2016** 5094126
- Federico S, Avolio E, Petracca M, Panegrossi G, Sanò P, Casella D and Dietrich S 2014 Simulating lightning into the RAMS model: implementation and preliminary results *Nat. Hazards Earth Syst. Sci.* **14** 2933–50
- Herrera A M, Suhandri H F, Realini E, Reguzzoni M and de Lacy M C 2016 goGPS: open-source MATLAB software *GPS Solut.* **20** 595–603
- Hofmann-Wellenhof B, Lichtenegger H and Wasle E 2007 *GNSS—Global Navigation Satellite Systems: GPS, GLONASS, Galileo, and More* (Berlin: Springer) pp 1–518
- Hong S Y and Lim J J O 2006 The WRF single-moment 6-class microphysics scheme (WSM6) *J. Korean Meteorol. Soc.* **42** 129–51
- Kalnay E 2003 *Atmospheric Modeling, Data Assimilation and Predictability* (Cambridge: Cambridge University Press) p 341
- Leick A, Rapoport L and Tatarnikov D 2015 *GPS Satellite Surveying* (New York: Wiley) pp 1–840
- Lindskog M, Ridal M, Thorsteinsson S and Ning T 2017 Data assimilation of GNSS Zenith total delays from a Nordic processing centre *Atmos. Chem. Phys.* **17** 13983–98
- Lorenc A C 1986 Analysis methods for numerical weather prediction *Q. J. R. Meteorol. Soc.* **112** 1177–94
- Mellor G and Yamada T 1982 Development of a turbulence closure model for geophysical fluid problems *Rev. Geophys. Space Phys.* **20** 851–75
- Molinari J and Corsetti T 1985 Incorporation of cloud-scale and mesoscale down-drafts into a cumulus parametrization: results of one and three-dimensional integrations *Mon. Weather Rev.* **113** 485–501
- Parrish D F and Derber J C 1992 The National Meteorological Center's Spectral Statistical Interpolation analysis system *Mon. Weather Rev.* **120** 1747–63
- Poli P, Moll P, Rabier F, Desroziers G, Chapnik B, Berre L, Healy S B, Andersson E and El Guelai F-Z 2007 Forecast impact studies of Zenith total delay data from European near real-time GPS stations in Meteo France 4DVAR *J. Geophys. Res.* **112** D06114
- Press W H, Teukolsky S A, Vetterling W T and Flannery B P 1992 *Numerical Recipes in C* (Cambridge: Cambridge University Press) pp 1–925
- Rawlins R, Ballard S P, Bovis K J, Clayton A M, Li D M, Inverarity G W, Lorenc A C and Payne T J 2007 The Met Office global four-dimensional variational data assimilation scheme *Q. J. R. Meteorol. Soc.* **133** 347–62
- Rohm W and Bosy J 2011 The verification of GNSS tropospheric tomography model in a mountainous area *Adv. Space Res.* **47** 1721–30
- Sánchez-Arriola J, Lindskog M, Thorsteinsson S and Bojarova J 2016 Variational bias correction of GNSS ZTD in the HAR-MONIE modeling system *J. Appl. Meteorol. Climatol.* **55** 1259–76
- Schaer S and Société helvétique des sciences naturelles. Commission géodésique 1999 *Mapping and Predicting the Earth's Ionosphere using the Global Positioning System* vol 59 (Zurich: Institut für Geodäsie und Photogrammetrie, Eidg. Technische Hochschule Zürich)
- Smagorinsky J 1963 General circulation experiments with the primitive equations. Part I, the basic experiment *Mon. Weather Rev.* **91** 99–164
- Takasu T and Yasuda A 2009 Development of the low-cost RTK-GPS receiver with an open source program package RTKLIB

- Int. Symp. on GPS/GNSS* (Jeju Korea: International Convention Center) pp 4–6
- Vedel H and Huang X-Y 2004 Impact of ground-based GPS data on numerical weather prediction *J. Meteorol. Soc. Japan* **82** 459–72
- Walko R L *et al* 2000 Coupled atmosphere-biosphere-hydrology models for environmental prediction *J. Appl. Meteorol.* **39** 931–44
- Yan X, Ducrocq V, Poli P, Hakam M, Jaubert G and Walpersdorf A 2009 Impact of GPS Zenith delay assimilation on convective-scale prediction of Mediterranean heavy rainfall *J. Geophys. Res.* **114** D03104
- Zou X, Deng Z, Ge M, Dick G, Jiang W and Liu J 2010 GPS data processing of networks with mixed single-and dual-frequency receivers for deformation monitoring *Adv. Space Res.* **46** 130–5
- Zumberge J F, Heflin M B, Jefferson D C, Watkins M M and Webb F H 1997 Precise point positioning for the efficient and robust analysis of GPS data from large networks *J. Geophys. Res.* **102** 5005–17