Ionospheric imaging in Africa

Alex T. Chartier,1,2 Joe Kinrade,1 Cathryn N. Mitchell,1 Julian A. R. Rose,1 David R. Jackson,1,2 Pierre Cilliers,3 John-Bosco Habarulema,3 Zama Katamzi,3 Lee-Anne Mckinnell,3 Tshimangadzo Matamba,3 Ben Opperman,3 Nicholas Ssessanga,3 Nigussie Mezgebe Giday,3 Vumile Tyalimpi,3 Giorgiana De Franceschi,4 Vincenzo Romano,4 Carlo Scotto,4 Riccardo Notarpietro,5 Fabio Dovis,5 Eugene Avenant,6 Richard Wonnacott,7 Elijah Oyeyemi,8 Ayman Mahrous,9 Gizaw Mengistu Tsidu,10 Harvey Lekamisy,11 Joseph Oku Owendo,12 Patrick Sibanda,13 Tsegaye Kassa Gogie,14 Babatunde Rabiu,15 Kees De Jong,16 and Adekola Adewale17

Received 28 May 2013; revised 19 November 2013; accepted 27 November 2013; published 8 January 2014.

[1] Accurate ionospheric specification is necessary for improving human activities such as radar detection, navigation, and Earth observation. This is of particular importance in Africa, where strong plasma density gradients exist due to the equatorial ionization anomaly. In this paper the accuracy of three-dimensional ionospheric images is assessed over a 2 week test period (2–16 December 2012). These images are produced using differential Global Positioning System (GPS) slant total electron content observations and a time-dependent tomography algorithm. The test period is selected to coincide with a period of increased GPS data availability from the African Geodetic Reference Frame (AFREF) project. A simulation approach that includes the addition of realistic errors is employed in order to provide a ground truth. Results show that the inclusion of observations from the AFREF archive significantly reduces ionospheric specification errors across the African sector, especially in regions that are poorly served by the permanent network of GPS receivers. The permanent network could be improved by adding extra sites and by reducing the number of service outages that affect the existing sites.


1. Introduction

[2] Human activities such as radar detection and satellite positioning are affected by ionospheric electron densities. Ionospheric imaging could be used to estimate the effects on these activities in the African sector. The work presented here utilizes a simulation approach to determine the accuracy of tomographic images of the ionosphere. This approach allows application developers to quantify the benefits of including observations from sites that are not currently operational. Real images are also presented and analyzed.

1.1. Ionospheric Tomography

[3] Global Positioning System (GPS) ionospheric tomography techniques invert observations of relative or calibrated slant total electron content (TEC) from dual-frequency GPS receivers to produce three-dimensional, time-dependent images of electron density. The technique used here, known as the Multi-Instrument Data Analysis System (MIDAS), is based on an algorithm by Mitchell and Spencer [2003] and developed by Spencer and Mitchell [2007]. The current version of the algorithm is described in detail in Chartier et al. [2012a].
1.2. Ionospheric Observations

MIDAS solves for three-dimensional electron densities at multiple times from differential phase measurements of slant TEC using a vertical basis function decomposition and a horizontal smoothing function. When relative slant TEC observations (from differential phase measurements) are used, there is no need to calculate the hardware biases because those biases only affect pseudorange measurements of TEC. Allain and Mitchell [2009] showed that a real-time version of MIDAS could improve single-frequency GPS position estimates by up to 25 m at midlatitudes during solar maximum. Alternative ionospheric imaging techniques have been developed by Bust et al. [2000, 2004], Schunk et al. [2004], Mandrake et al. [2005], and Angling and Jackson-Booth [2011]. Andreeva et al. [2000] studied the equatorial ionization anomaly using an ionospheric tomography algorithm.

Figure 1. The map shows working dual-frequency GPS receiver sites on the IGS network in blue and additional sites from the AFREF and UNAVCO networks in red. Only sites that produced usable observations in the period 2–16 December 2012 are shown here.

1.3. Ionospheric Models

The work described in this paper relies on realistic simulations of ionospheric electron density to serve as a ground truth. The latest version of International Reference Ionosphere (IRI), IRI-2012, is chosen for this purpose. The observational noise that is to be expected from subgrid-scale structures is dealt with separately (see section 2.3). IRI-2012 is an empirical model of the ionosphere based on a wide range of ground and space data, including incoherent scatter radars and topside sounders [Bilitza et al., 2011]. It is the result of collaboration between the Committee on Space Research and the International Union of Radio Science that began in 1969. IRI estimates the monthly median electron density, ionized gas composition, and temperature in the altitude range 50–1500 km.

2. Method

The quality of ionospheric reconstructions is difficult to assess without an independent ground truth. In this experiment, we use a modeled ionosphere from IRI-2012 as a ground truth. Simulated GPS TEC observations through this modeled ionosphere are created and used in MIDAS inversions. As well as allowing for comparison of the images with the ground, this approach allows us to quantify the benefits of including observations from additional sites that are not currently operationally available. A similar approach was used by Dear and Mitchell [2006] in Europe and by Zapfe et al. [2006] in South America. Differences between the model and the reconstructed images are then due to a lack of observations or due to poor assumptions about the nature of the solution. While the MIDAS inversion technique can use IRI to create basis functions and to provide an initial guess at the solution, those options are disabled for this experiment. Instead, we use two basis functions derived from Chapman’s equations and a 4° horizontal grid. In addition to assuming that the ionosphere can be adequately represented by these constraints, a regularization condition is applied that favors solutions with zero second derivative in each horizontal direction and in time. These assumptions are necessary to obtain a unique solution and to cover data gaps, but it is noted that the assumptions could prove problematic if they are not appropriate for a specific ionosphere. The selection of IRI-2012 as the “truth” ionosphere might artificially enhance the performance of the imaging algorithm because IRI-2012 is smoother than the real ionosphere, and our algorithm favors smooth solutions. This problem is addressed by adding realistic noise caused by subgrid-scale structures to the observations. However, scintillation is not modeled here.

2.1. Simulating the Ionosphere

IRI-2012 is used to create three-dimensional fields of electron density values on the same 4° grid that will be used for the reconstructions. MIDAS uses observations from Africa, but it also recognizes other applications of GNSS signals, such as for ionospheric imaging. By including data from these receivers in simulated inversions, it is possible to assess the impact of observations that come from geographically feasible locations. A comparison can be made between the quality of ionospheric reconstructions from existing data and the improvements that can arise from including additional receiver stations.
a time window around the inversion, so it is necessary to simulate the ionosphere for multiple times. In this case, we use a time window of 7 h and 30 min with a time step of 30 min. The result is that for each inversion, 15 IRI simulations are created at 30 min intervals. The resulting IRI electron density values are arranged into a simulated state vector, $x_{IRI}$.

2.2. Simulating a Receiver Network

In order to find the upper limit of imaging accuracy possible using the chosen grid for the MIDAS algorithm, it is necessary to simulate a network of receivers that would provide adequate observation coverage. This is achieved by creating a regularly spaced list of coordinates to represent fictitious receivers at 8° intervals in latitude and longitude. This receiver spacing was chosen because it was found that increasing receiver coverage above 8° spacing made almost no difference to image accuracy. This list of coordinates is combined with the known position of the GPS satellites to find the trajectories of the rays that would be observed by the simulated receiver network. Real receiver networks can be also used in the simulation approach, with the added advantage that data outages can be taken account of. This is described in the next section.

2.3. Simulating Observations

As already noted, it is necessary to create TEC “observations” of the simulated ionosphere in order to produce reconstructed images. This is achieved using an observation operator, $H$, that is based on the trajectories of real or simulated observations. $H$ describes the raypath contributions of the observations to the grid. $H$ is created by tracing raypaths from the GPS satellites to the receivers at the times when data are received. In order to create a vector of simulated observations, $z_{IRI}$, we multiply the simulated state vector by the observation operator, i.e.,

$$z_{IRI} = Hx_{IRI}$$  \hspace{1cm} (1)

Figure 2. (left) An IRI modeled truth. (right) A reconstructed image based on observations of the modeled truth from a fictitious receiver network (shown in white). The model and reconstructed image are from 12:00 UT on 7 December 2012.

Figure 3. The RMS errors of reconstructions based on simulated observations from a dense network of simulated receivers are shown here. The simulated receiver sites are shown in white. The IRI simulations that the observations are based on are used as the ground truth here. The RMS errors are based on hourly reconstructions from the 2 week test period (2–16 December 2012).
We assume that real GPS differential phase observations of slant TEC do not contain significant errors, although they do contain cycle slips. It is possible that the observations are not representative of grid-scale structures—there could be significant subgrid-level “noise” in the real ionosphere. This could happen if structures exist in the ionosphere that are too small to image using the specified resolution. Our simulated ionosphere, IRI-2012, will not have these subgrid-scale structures because it is defined on the same grid that will be used in the inversion. The simulated ionosphere is also far smoother than the real ionosphere. It is important to include realistic errors of representativeness in the simulation so that the inversion accuracy is not artificially enhanced. This is achieved by creating images, $x_{\text{REAL}}$, of the real electron density distribution using real observations of slant TEC, $z_{\text{REAL}}$, and then calculating the residuals, $r$, of the observations from the images, i.e.,

$$r = z_{\text{REAL}} - Hx_{\text{REAL}}$$  \hspace{1cm} (2)

These residuals are added to the simulated observations of slant TEC in order to take account of the effects of subgrid-level structures on image accuracy.

### 2.4. Reconstructions Using Simulated Observations

The simulated observations are inverted in order to reconstruct the simulated ionosphere. The normal MIDAS inversion procedure is followed. First, the problem is mapped from electron density space to basis function space. A mapping function, $M$, is used to make the transformation. Then a regularization condition, $R$, is applied that penalizes solutions that contain nonzero second derivatives of electron density in horizontal space and in time. There is no regularization of the vertical profile. In practice, a weighting term, $\lambda$, has to be included in order to balance the effects of the measurements and the regularization on the solution. Within MIDAS, a heuristic choice is made to define the regularization weighting as

$$\lambda = \text{trace}(M^T H^T H M)/\text{trace}(R)$$  \hspace{1cm} (3)

Finally, the constrained problem is inverted to obtain a solution, $x_{\text{RETRIEVED}}$, for all the times in the time window, i.e.,

$$x_{\text{RETRIEVED}} = (M^T H^T H M + \lambda R)^{-1} M^T H^T z_{\text{IRI}}$$  \hspace{1cm} (4)

The central slice in time is selected as the final image. The retrieved solution, $x_{\text{RETRIEVED}}$, is compared with the original ionospheric simulation, $x_{\text{IRI}}$, in order to determine the accuracy of the imaging technique, given the available data.

### 3. Results

Following the procedure described in section 2, three sets of simulated inversions were produced for the maximum AFREF data availability period (2–16 December 2012). Differences between the two sets of images show the improvements in accuracy that can be achieved by using additional receivers.
first set of images is based on the simulated, regularly spaced receiver network. These results provide an estimate of the upper limit of imaging accuracy achievable using the MIDAS algorithm at a 4° grid resolution under ideal conditions. The other two sets of images are created to determine the imaging accuracy achievable using real GPS receiver networks. One set of images was based on simulated observations at the locations of the IGS receivers, while the other set was based on all the available receivers. In each case, images were produced every 30 min throughout the test period.

### 3.1. Imaging Under Optimal Conditions

[18] Although insufficient observation coverage is likely to be the primary source of error in ionospheric images, it is possible that inherent properties of the imaging technique also limit accuracy. The results presented in this section demonstrate the performance of the imaging technique when provided with high-density (8° spaced) and uniform GPS receiver coverage. One hundred receivers are used in total. An example of the model truth and the image obtained from this high-density simulated network is shown in Figure 2.

[19] Images such as the one in Figure 2 are produced at 30 min intervals over the period 2 – 16 December 2012. In order to assess the errors of the images, differences between the reconstructed images and the modeled truth are calculated over the whole period. The root-mean-square (RMS) errors of the images from the fictitious receiver network are shown in Figure 3.

[20] The results in Figure 3 show that relatively small errors can be achieved when a dense network of receivers is available—errors range from 0 to 5 TECU (TEC unit, 1 TECU = 10^{16} \text{ el m}^{-2}) here. Errors are clearly highest at the locations of the two bands of increased ionization created by the Appleton anomaly (around 10°S–0° and 15°N–20°N).

---

**Figure 5.** (left) The IRI simulations, (middle) the reconstructions based on all the available data, and (right) the reconstructions based on just the IGS data. (a) For 22:00 UT on 2 December 2012. (b) For 17:00 UT on 3 December 2012. (c) For 12:00 UT on 7 December 2012. The GPS receiver sites used to make each set of reconstructions are shown in white.
As well as the higher TEC values present here, the TEC gradients seen in this region are likely to contribute to these higher errors. The imaging technique includes a regularization condition that favors zero second derivatives in the horizontal directions—a condition that is clearly broken as we move over the Appleton anomaly in a longitudinal direction.

### 3.2. Data Coverage

[21] In order to illustrate the typical data coverage available for each set of images, a pair of case studies are shown in Figure 4. The 300 km pierce points of all the rays available for use in each image are shown. These rays are collected at 30 min intervals over a 7.5 h time window. The case study shows data from 00:00 to 07:30 UT on 3 December 2012.

[22] Figure 4 shows that the full network provides far more raypath coverage than the IGS network alone. The IGS network has only isolated patches of coverage during the test period, while the full network only has a few large gaps. There are numerous redundant receivers in the full network—a similar level of coverage could be achieved with far fewer receivers, but such a network would be vulnerable to station outages.

### 3.3. Case Studies

[23] It is useful to examine the images from the two simulations alongside the original IRI simulations in order to understand how the structures in the ionosphere affect the resulting image accuracy. Figure 5 shows a series of case studies that depict the way the image quality varies depending on the ionospheric state. The examples were selected to show a range of different features that occur at different times of day.

[24] All the case studies shown in the right column in Figure 5, based on just IGS data, overestimate the IRI simulated truth in the left column more than the images based on all the data. However, the errors are generally overestimates in both cases. A nighttime case study is not included here because the lack of large-scale ionospheric structuring and lower electron densities at night make imaging much easier.

[25] The different case studies give an insight into the reasons for the overestimation. Figure 5a shows a significant overestimation of TEC in the western region of the IGS reconstruction. This is caused by the regularization condition, which extrapolates the gradients observed in the north and east of the image. The lack of data in the western region allows the TEC values to continue increasing until the edge of the image. This does not occur in the reconstruction based on all the available data because there are numerous active receiver sites in the western part of the image. The lack of data could equally have resulted in artificially low TEC values, but the distribution of receiver sites in this case means that TEC generally increases from the observed to the unobserved parts of the grid. This positive gradient causes the artificial enhancements observed.

[26] The IGS-only reconstruction in Figure 5b has the western TEC enhancement most clearly shown in Figure 5a, but Figure 5b also contains a more unusual artifact. In this case, the IGS-only reconstruction has underestimated the northern band of ionization caused by the equatorial ionization anomaly in the northeastern sector. It appears that the sites above and below this phenomenon have measured only small positive gradients toward the band of ionization, and therefore, the interpolation has resulted in an underestimate. The reconstruction with all the available data shows that it is possible to image this phenomenon accurately when two east African receivers are present.

[27] In Figure 5c and, to a lesser extent, Figure 5b, both reconstructed images overestimate the TEC values of the IRI truth image. The overestimation is caused by the regularization condition, which extrapolates a constant gradient across data-sparse regions. This is evident in the west of the images, where there are few observations available.

### 3.4. RMS Errors

[28] To measure the accuracy of the images, we calculate the differences between the vertical TEC from the images and the vertical TEC from the simulations that the images are based on. Spatially distributed root-mean-square (RMS) errors are calculated based on the errors of the entire period (2–16 December 2012) and plotted in Figure 6.

[29] The results show that the reconstructions based on more data have far lower errors. The overall RMS error figures are 3 TECU for the fictional, ideal network, 4.5 TECU for the “full” network, and 9.5 TECU for the IGS-only network. The errors are generally larger farther from the receiver sites, as
shown in Figure 6. The IGS-only images had very large RMS errors (around 15 TECU RMS) in the region south of 15°N and west of 15°E. This is because there were no observations from IGS receivers in this region during the test period.

3.5. Images Based on Real Observations

The results of sections 3.3 and 3.4 showed that it is possible to get reasonably accurate images of the ionosphere over the mainland area of Africa with the full (IGS, AREF, and UNAVCO) receiver network using a simulation approach. In this section, images based on real data are shown. The case studies selected here match the times of the simulated images described in section 3.3. The images based on real measurements of slant TEC are shown in Figure 7.

[31] The images in Figure 7 correspond to the case studies shown in Figure 5. In general, TEC values are lower in Figure 7. This is not surprising—the IRI simulated ionosphere is only intended to provide a monthly median specification. The images in Figure 7 also appear to have some of the same artifacts described in section 3.4. The images based on only IGS data in Figures 7a and 7b have large western TEC enhancements because of lack of data. The images based on all the data in Figures 7a and 7b clearly show the two latitudinal bands of ionization associated with the

Figure 7. (left) Reconstructions based on all the available data. (right) Reconstructions based on just the IGS data. (a) For 22:00 UT on 2 December 2012. (b) For 17:00 UT on 3 December 2012. (c) For 12:00 UT on 7 December 2012.
equatorial ionization anomaly, a feature that is absent from the IGS-only images. Figure 5c shows a clear overestimation of TEC values compared with IRI’s truth. In the absence of an independent truth, it is impossible to say whether the images in Figure 7c overestimate or underestimate the true TEC values present in the ionosphere. Figure 7c does show that it is possible to produce reasonably good images from just the IGS receiver data—both images have similar features and absolute values across the land area. In this case, the real ionospheric conditions matched the inversion regularization condition closely enough that it was possible to extrapolate the available data accurately.

4. Discussion and Conclusions

[32] Accurate ionospheric specification is necessary for improving satellite positioning and radar detection. The aim of this study was to determine how much GPS receiver coverage is required to accurately specify the ionosphere over Africa. Figure 6, presented in section 3.4, shows the RMS errors of two sets of images, which are based on an IRI simulated truth ionosphere. The results show that it is possible to achieve below 5 TECU RMS error over most of Africa if data from the full network are present. The full network is the combined output of the IGS, UNAVCO, and AFREF stations. This RMS error estimate takes into account imaging problems caused by subgrid-scale structures (see section 2.3) but assumes that real large-scale ionospheric structures are similar to those in IRI-2012. Most of the full network stations are from a short-term campaign, so the IGS-only results represent the accuracy that can currently be achieved. In this case, RMS errors are often well above 10 TECU over Africa. The situation is even worse over the ocean, where RMS errors from the IGS-only simulations exceed 20 TECU. It should be noted that the presence of small-scale structures, which are known to cause GPS signal scintillation, could have a significant negative impact on image accuracy. The phenomenon of scintillation was not considered in this study. The results shown here make a clear case for the addition of more permanent operational receiver sites in Africa.

[33] While there is a clear need for more receiver coverage, care must be taken in siting the receivers to maximize their effectiveness. Figure 4 shows the raypath coverage obtained from the two networks. In the full network, numerous receivers are grouped together. The observations from these clustered receivers provide little more ionospheric information than would be provided by a single receiver in the center of the cluster. The same number of receivers could specify the ionosphere far more accurately if the receivers were evenly distributed across the grid. Alternatively, the same image quality could be achieved using a far smaller number of receivers. However, there is an important benefit to having clusters of receivers that is not obvious from these results. The GPS receiver sites frequently experience extended data gaps—in fact, the IGS network has a number of receivers that produced no usable data during the test period. The creation of multiple receivers in close proximity to each other should be encouraged in areas where accurate ionospheric specification is required.

[34] When considering the potential deployment of new GPS receivers for the purpose of ionospheric specification, it is useful to have an idea of the maximum achievable accuracy possible through ground-based GPS tomography. The results presented in section 3.1 give an estimate of this upper limit by producing images from a dense, fictitious network of receivers. The RMS errors of these images (shown in Figure 3) are lower than those presented in Figure 6 in section 3.4, which is as expected since those results are based on limited and unevenly distributed observations. However, it should not be assumed that perfect images could be produced only if we had enough GPS receivers. As shown in Figure 3, RMS errors of up to 5 TECU are still present in images based on a dense network. These errors are partly due to the inaccuracy of the assumptions used in the imaging process. It is not always possible to reproduce the true ionosphere by a linear combination of two vertical basis functions, and the ionosphere does not always match the specified regularization condition. For example, Zapfe et al. (2006) achieved optimal image accuracy by using four vertical basis functions. In general, a greater number of basis functions provide the inversion with more degrees of freedom. This might improve the image accuracy if there are enough observations to constrain the solution but could otherwise lead to instability.

[35] Different GNSS systems may be also utilized to improve ionospheric image accuracy. In situations where there are less usable satellites than usable receivers, as is the case with nearly all ionospheric imaging, image quality will normally benefit more from the addition of satellites than from the addition of receivers. This is because the new satellites will create more new raypaths (one to each visible receiver) than a new receiver would (one to each visible satellite). Bearing this in mind, any new receivers installed should be designed to receive data from as many different GNSS networks as possible. Receivers placed in coverage gaps will clearly improve imaging accuracy far more than those placed close to existing receivers. The inclusion of observations that provide detailed information on the vertical electron density profile, such as observations from ionosondes or radio occultation measurements, would also provide significant improvements to ionospheric imaging accuracy, as shown in Chartier et al. (2012b).

[36] In any realistic situation, the number and type of observations are likely to be insufficient to specify the ionosphere without making any assumptions. For this reason, it is necessary to use techniques such as regularization and basis function transformations. In section 3.3, several image artifacts were highlighted and attributed to the regularization condition used here (which favors zero second derivatives in the horizontal directions and in time). The most pronounced error was the TEC enhancement in the southwest of the images. This problem was caused by the lack of data in that region. It would be possible to overcome this problem by using different imaging techniques. For example, a background model term could be included in the inversion. A background model was not used here because such an approach could introduce significant biases across the image.

[37] The results in section 3.5 show that it is sometimes possible to produce reasonably good images from very few observations. In particular, Figure 7c shows very similar images produced from the IGS and full networks. This was possible because the inversion assumptions (regularization and basis functions) matched the real ionospheric conditions quite well at that time. Figures 7a and 7b show that this is often not the case—at least the inversion assumptions are seldom sufficient to provide good image accuracy in the absence of sufficient data.
A future study should explore the relationship between ionospheric image accuracy and GPS positioning accuracy. There is a direct relationship between the line-of-sight signal delay and the slant TEC along a satellite-to-receiver path, but the positioning errors experienced by users depend on the number and location of available satellites as well as the TEC present. Allain and Mitchell [2009] presented a method for calculating positioning error from maps of electron density that could be applied to the maps presented in this paper. Future work should also seek to translate these scientific results into metrics that are directly useful for other applications.

In summary, the results of this paper show that African ionospheric images can be made significantly more accurate if additional receivers (beyond those available through the IGS network) are used. The use of these additional receivers can reduce RMS TEC errors by a factor of 2 over a large part of Africa. The results suggest that any new operational receivers should be deployed far from currently operating receivers. Efforts should be also made to increase the reliability of the existing network so that network redundancy is not required. Both new and existing receivers should be made capable of using as many different GNSS networks as possible.

Acknowledgments. The authors acknowledge the support of the European Space Agency in conducting this work. The provision of ionospheric observations by the IGS, UNAVCO, and AFREF networks is acknowledged. C.N.M. acknowledges support from the Royal Society Wolfson Fund. This work was supported by the Systems Centre, the EPSRC-funded Industrial Doctorate Centre in Systems (grant EP/G037353/1), the ESA Alcantara initiative (contract 4000106579/12/F/MOS), and the Met Office. The authors acknowledge the work of Rich Pawlowicz in developing the M_Map plotting tool used here.

References

Allain, D. J., and C. N. Mitchell (2009), Ionospheric delay corrections for single-frequency GPS receivers over Europe using tomographic mapping, GPS Solutions, 13(2), 141–151.


