Near real-time input to a propagation model for nowcasting of HF communications with aircraft on polar routes


1Department of Engineering, University of Leicester, Leicester, UK, 2Department of Radio Physics, Faculty of Physics, Saint Petersburg State University, Saint Petersburg, Russia, 3Department of Physics, Lancaster University, Lancaster, UK, 4Natural Resources Canada, Ottawa, Canada

Abstract

There is a need for improved techniques for nowcasting and forecasting (over several hours) HF propagation at northerly latitudes to support airlines operating over the increasingly popular trans-polar routes. In this paper the assimilation of real-time measurements into a propagation model developed by the authors is described, including ionosonde measurements and total electron content (TEC) measurements to define the main parameters of the ionosphere. The effects of D region absorption in the polar cap and auroral regions are integrated with the model through satellite measurements of the solar protons (>1 MeV) and the X-ray flux in the 0.1–0.8 nm band, and ground-based magnetometer measurements which form the Kp and Dst indices of geomagnetic activity. The model incorporates various features (e.g., convecting patches of enhanced plasma density) of the polar ionosphere that are, in particular, responsible for off-great circle propagation and lead to propagation at times and frequencies not expected from on-great circle propagation alone. The model development is supported by the collection of HF propagation measurements over several paths within the polar cap, crossing the auroral oval, and along the midlatitude trough.

1. Introduction

Extensive HF propagation measurements have been made at northerly latitudes over a number of years by the University of Leicester and colleagues [see, e.g., Warrington et al., 1997; Zaalov et al., 2003; Rogers et al., 1997, 2003; Siddle et al., 2004a, 2004b]. Of particular relevance to this paper, measurements undertaken in the polar cap found that the presence of convecting patches and Sun-aligned arcs of enhanced electron density can lead to signals arriving in directions displaced from the great circle path by up to 100° [Warrington et al., 1997; Zaalov et al., 2003] and at times and frequencies not expected by great circle propagation alone. The measurements of direction of arrival undertaken in our experiments give insight into the complex propagation mechanisms present at high latitudes. It is particularly important to note that these propagation mechanisms have significant impact on the coverage of HF transmissions where the signals are reflected from the high-latitude ionosphere. It was also found that the signals can arrive at the receiver over a range of directions with, for example, azimuthal standard deviations of up to 35° at frequencies of 2.8, 4.0, and 4.7 MHz being observed on one path from Isfjord, Svalbard to Alert, Canada [Warrington, 1998]. Similar measurements have also been undertaken at auroral latitudes [Warrington et al., 2006].

Patches are formed in the dayside auroral oval [see, e.g., MacDougall and Jayachandran, 2007] during periods of southward directed interplanetary magnetic field (IMF) (Bz < 0) and the associated high levels of geomagnetic activity and generally convect in an antisunward direction across the polar cap into the nightside auroral oval, whereas arcs occur when geomagnetic activity is low and the IMF is directed northward (Bz > 0) and drift in a duskward direction [Buchau et al., 1983].

This type of propagation is exemplified by the measurements of the direction of arrival of 8.0 and 11.1 MHz signals received over the path from Qaanaaq, Greenland to Alert on 17/18 October 2014 presented in Figure 1. It is evident from these data that the signals frequently arrive from directions well displaced from the great circle direction, often without the presence of an on-great circle component. It is at times such as these that communications would be supported but not expected when only great circle propagation is considered.
2. Previous Ray Tracing Model

Zaalov et al. [2003, 2005] reported on a ray tracing model that can accurately reproduce many of the direction of arrival features observed in experimental measurements. Simulations making use of the numerical ray tracing code developed by Jones and Stephenson [1975] are employed to estimate the raypaths from a transmitter location through a model ionosphere. Initially, a background ionospheric model is produced, which is then perturbed to include the various ionospheric features (in particular, patches, arcs, auroral zone irregularities, and the midlatitude trough) that are expected to significantly affect the propagation of the radio signals.

Some of the more pertinent points are outlined below, full details being reproduced elsewhere [Zaalov et al., 2003, 2005]:

1. The background ionosphere comprises two Chapman layers, the main parameters of which (critical frequency, critical height, and vertical-scale height of each layer) were determined from vertical ionospheric soundings. In view of the limited number of ionosondes available at high latitudes (and this number is decreasing), it was not possible to obtain snapshots of the ionospheric parameters sufficient to define the background ionosphere. Consequently, curves were fitted to the required parameters as a function of time for several ionosonde stations. These curves were then used as the basis of defining the latitudinal and longitudinal variation of the background ionosphere in terms of a series of spline fit curves, with longitudinal values obtained by rotating measurements along lines of constant geomagnetic latitude with appropriate time shifts of up to ±12 h.

2. Patches of enhanced electron density are modeled as an arbitrary number of Gaussian distributions with approximately equal longitudinal and latitudinal scale. The temporal evolution of the patches relative to the propagation path is simulated by means of a convection flow scheme coupled with the rotation of the Earth beneath the convection pattern, the precise form of which depends upon the components of the IMF [Lockwood, 1993].

3. Sun-aligned arcs are defined within the model by a small number of three-dimensional Gaussian perturbations in electron density of different spatial scales (altitude, longitude, and latitude) randomly distributed near to the center of the arc. Several Gaussian perturbations are combined in defining the shape of each modeled arc in order to prevent the shapes of the arcs being too stylized.

Figure 1. Measurements of direction of arrival of signals received over the Qaanaaq to Alert path at 8.0 and 11.1 MHz on 17–18 October 2014. The great circle direction is indicated by the dashed lines.
4. An analytical approximation to the trough model presented by Halcrow and Nisbet [1977] is employed. Their model is trapezoidal in form, whereas our model has a smooth variation in electron density perturbation that is more physically realistic and is in a form suitable for ray tracing.

5. The model also includes other features such as the plasma irregularities found in the auroral oval. In addition to simulating the raypaths of the HF radio waves, the effect of D region absorption is also incorporated into the model. There are three principal mechanisms that are included: diurnal absorption caused by solar UV [Davies, 1990], absorption associated with X-ray flux, and particle flux resulting from solar flares [Sauer and Wilkinson, 2008].

The area coverage to be expected from a transmitter at a given frequency, time, and location may then be estimated by ray tracing through the model ionospheres. A large number of rays are launched from the transmitter in an azimuth/elevation grid. Each ray is assigned a power dependent upon the transmitter power, the ray density at the transmitter, and the transmit antenna gain in the direction of the ray. The rays are then traced through the model ionosphere and the power adjusted to take into account the D region absorption by noting the vertical absorption at the point that the rays cross 90 km height and applying a secant correction for the angle of incidence. The signal strength at the receiver is then estimated by adding the power conveyed by the rays to the receive antenna.

An example outcome of this modeling process is presented in Figure 2, where the received signal strength is estimated over the polar region for a transmitter located at Qaanaaq, Greenland at frequencies of 8.0 and 11.1 MHz, corresponding to measurements presented in Figure 1. Signal strength is presented as S units, the scale commonly used on HF communications receivers with S1 being very weak (−121 dBm), stepping in 6 dB increments per S unit to S9 (−73 dBm) indicating a relatively strong signal and then in dB exceeding S9. Taking just the background ionosphere into account, reception is not expected at Alert as it is within the skip zone at both frequencies. Including a set of randomly located patches significantly alters the ground

![Figure 2](image-url)
coverage pattern, in particular resulting in signal coverage within the expected skip zone. In these cases, reflection has occurred from the patches rather than from the smooth background ionosphere and, consequently, the signal often arrives from directions offset from the great circle path, in agreement with the measurements presented in Figure 1.

In considering the effect of the presence of patches, it is important to remember that in reality the patches will be distributed differently from the model (we do not know exactly where they are), will evolve in time, and will move generally following a convection pattern. To estimate the effect of patches on a statistical basis, a large number of simulations are undertaken, and the median and decile signal strengths calculated.

3. Developments to the Ray Tracing Model

The methodology described in section 2 has served us well in modeling specific historical events and investigating the effect of the presence of various ionospheric features. However, building the ionospheric model does require significant manual input and is not readily adaptable to automated running as required in routine nowcasting and forecasting applications. We are currently developing the model for such applications using data that are available in near real time.

The approach that we are currently investigating is to start with the International Reference Ionosphere (IRI) [Bilitza, 1990; Bilitza et al., 2011, 2014] and to perturb it by adjusting the IG and RZ indices (the global ionospheric index and the sunspot number usually input as a 12 month running mean) to force the IRI output to match measurements made at a number of sites to form the basis of the background ionosphere model employed in the ray tracing procedures. The IG and RZ values are not expected to be the same over the entire region of interest; therefore, it will be necessary to adopt a mapping technique to give a smooth variation of the ionospheric parameters over the region. The modeled ionosphere will then be further perturbed to include features such as the convecting patches, the parameters of which will also be informed by measurements. One problem is the high variability of the high-latitude ionosphere and the relative scarcity of real-time ionospheric measurements over the region.

Ionosonde measurements are perhaps the first to come to mind when considering HF propagation problems. Galkin et al. [2012], for example, have integrated such measurements into a real-time IRI model. While there are a significant number of ionosondes worldwide, coverage is not uniform, and there are no instruments over the oceans. Of particular relevance from the point of view of this paper, coverage is sparse at high latitudes, and the number of high-latitude ionosondes is decreasing with the possible exception of the Canadian High Arctic Ionospheric Network [Jayachandran et al., 2009]. Furthermore, high-latitude ionograms are not always easy to interpret, either manually or automatically. The signatures of various high-latitude ionospheric features on vertical ionograms have been considered by Moskalova and Zaalov [2013]. Two typical ionograms from Tromsø, Norway are given in Figure 3, the right panel corresponding to a time included in Figure 1, and...
the left panel to 12 h earlier. As evidenced by this figure, at times it is relatively easy to obtain the required parameters \((f_2, h_m, B_0, \ldots)\) from the ionogram, whereas at other times the required features cannot be identified and the autoscaling fails to produce sensible results. Oblique ionospheric measurements, perhaps including directional information, could also form a valuable input. However, the installation of such a network to support HF nowcasting in the polar regions is extremely unlikely, and currently, we are not pursuing this as a source of data to drive the model. We have used directional information to validate the model, but the measurement equipment is limited in geographical coverage and is not intended for permanent installation.

Further sources of ionospheric measurements that may be used are networks of GPS receivers (in particular, the International GNSS Service (IGS) network [Dow et al., 2009]) capable of measuring the total electron content along paths between individual receivers and individual satellites (this is referred to as the slant TEC or sTEC). Many users have then converted the sTEC values into estimates of the vertical TEC (evTEC), making simplifying assumptions about the ionosphere, which we do not intend to do. Previous workers, for example Komjathy et al. [1998] and Hernandez-Pajares et al. [2002], have used GPS TEC measurements to update the IRI concentrating on the TEC values. Although the IRI on its own cannot match the day-to-day variations seen in measurements and exhibits anomalies at high latitudes [Figurski and Wielgosz, 2002], it represents a useful mapping technique to create a smooth complete model ionosphere from a discrete set observations. Lack of coverage over the oceans and polar region is a large source of error, but long-term studies [McNamara, 2009; McNamara and Wilkinson, 2009] show that significant correlation of \(f_2\) values exists between ionosonde stations separated by between 700 and 1500 km, depending on time of day and phase of solar cycle. These are similar to the correlation distance for TEC [Shim et al., 2008] and suggest that the IGS network of GPS stations has an adequate density for model input, at least in some areas.

Furthermore, the IRI allows the input (indirectly) of GPS based measurements of TEC and the output of \(f_2\) values. Maltese et al. [2012] have shown that these two parameters do not always respond in synchrony to various solar and geomagnetic influences, but that the IRI can provide a useful estimate of the equivalent slab thickness (the ratio of TEC to \(N_m F_2\)), to convert between the two. Barabashov et al. [2006] have compared various correction methods to improve the fit between the model predictions and measured values of \(f_2\) and found that applying a correction to the IRI's topside is most beneficial.

In addition to using GPS TEC measurements to provide an estimate of the background ionosphere, they can also be used to establish the presence of patches, estimate their location, and estimate their intensity. However, there are limitations since the GPS coverage is not complete; and therefore, patches might be present but not observed and hence accurately establishing the total number of patches present is unlikely to be possible. However, the model can be run several times with different numbers of patches, positions, and intensities guided by the measurements to give a range of possible outcomes to produce a statistical prediction.

Slant TEC (sTEC) measurements made at Alert with a number of GPS satellites at the same time as the measurements of Figure 1 are presented in Figure 4 alongside measurements made at a midlatitude site at a similar longitude (Algonquin, Canada). The difference between the nature of the measurements made at the two sites is striking: at the midlatitude site the traces are (more or less) smoothly varying, whereas at the high-latitude site, significant deviations are evident due to increases in sTEC resulting primarily from the presence of the patches.

The problem we face is to determine IG and RZ values from the measured sTEC values by successive comparison with IRI-predicted values and to then use these values in the IRI to give estimates of the relevant parameters for the ray tracing model, namely, \(f_2, h_m, F_2, B_0\), and the bottomside thickness parameter \(B_0\). On a historical basis, the monthly mean values of \(IG\) and \(RZ\) are closely related (see Figure 5), and in the absence of sufficient information to treat these two parameters independently, their relationship has been fixed by a curve fitted to the data.

Examples of using the sTEC measurements to give revised estimates of \(f_2\) through the above process are given in Figures 6 and 7 for a Canadian site (Alpena) and a UK site (Fairford), respectively, which also show \(f_2\) values measured by ionosondes at these locations together with the monthly median IRI \(f_2\) values. For Alpena, the IG and RZ values were obtained from sTEC measurements made at Algonquin, Canada with
the satellite signals selected to include only those where the 350 km penetration points were within 500 km of the Alpena ionosonde site (444 km station separation). For Fairford, the sTEC measurements were made at Hailsham, UK, and the satellite signals selected to include only those where the 350 km penetration points were within 500 km of the Fairford ionosonde site (155 km station separation). It is evident from these figures that the $f_0F_2$ values obtained using the sTEC-derived values of $IG$ and $RZ$ to drive the IRI have a closer agreement with the measured values than the monthly median IRI values during the day, at times closely following deviations of several MHz from the IRI median values. At night, however, the TEC-driven IRI values do not follow the measured values as well as during the day, and sometimes the change in $f_0F_2$ is in the wrong sense (i.e., an increase in $f_0F_2$ is generated when a decrease is required, and vice versa).

Figure 8a shows the $IG$ and $RZ$ values obtained from the sTEC measurements made at Algonquin for the 17 October 2014, and Figure 8b shows $f_0F_2$ derived from the IRI using these $IG$ and $RZ$ values compared with ionosonde measurements of the same parameter made at Alpena. Good agreement is obtained on that day, although it is evident that the $f_0F_2$ values are close to the monthly median values obtained from the IRI. For comparison, the $f_0F_2$ values for the 24 October 2014 are presented in Figure 8d, and the corresponding $IG$ and $RZ$ values in Figure 8c. For this second day, the measured sTEC and $f_0F_2$ values differ markedly from the monthly medians during the daytime.
and it is evident in this case that the sTEC-derived \( f_{\text{F2}} \) values are a better estimate of the measured \( f_{\text{F2}} \) than the median values. An example for two European stations, Hailsham (GPS) and Fairford (ionosonde), is presented in Figure 9. It is important to note that the \( IG \) and \( RZ \) values derived from the TEC measurements differ between the Canadian and UK sites (as indeed they do between much closer sites), and it will be necessary to establish appropriate correlation distances (likely to be of the order of 100 s of kilometers to around 1500 km [see, e.g., McNamara, 2009 and McNamara and Wilkinson, 2009]) to incorporate into mapping techniques to obtain the geographic variation of the required ionospheric parameters over large areas of the Earth.

Figure 10a shows the occurrence probability (expressed as a percentage) in 3 h periods that the TEC-driven IRI \( f_{\text{F2}} \) values for Alpena are either an improvement on the monthly median IRI values or are classed as unchanged (i.e., are within 0.2 MHz, a somewhat arbitrary value but broadly corresponding to the precision with which \( f_{\text{F2}} \) can be obtained from the ionograms). It is clear from this figure that a significant number of improvements occur during the daytime. The magnitude of the improvement is indicated in Figure 10b of this figure where the RMS error is around 0.5 MHz throughout the day for the sTEC-driven IRI compared to a peak of around 1.8 MHz for the monthly IRI values. Figures 10c and 10d of the figure show the same...
information for Fairford. Again, a significant improvement is evident for daytime, but in this case the RMS error increases significantly after midnight local time. The reason for this is currently unclear but may be related to nighttime B0 or topside electron density values in the IRI.

Both of the comparisons made in Figures 6 and 7 are for midlatitude stations where the ionosphere is relatively benign. As noted previously (Figure 4), TEC values are significantly perturbed by the presence of large-scale ionospheric irregularities (e.g., patches), increasing the measured values above the background level. At high latitudes, therefore, we have adopted a technique of manually observing the sTEC traces to estimate the background levels based on the variation in the previous few hours.

Figure 8. (a and c) The values of IG and RZ required to give good agreement between measured sTEC and IRI-derived sTEC for the GPS receiver located at Algonquin, Canada. (b and d) The foF2 values obtained from the IRI with the sTEC-derived IG and RZ values compared with the measured foF2 values at Alpena, Canada. Figures 8a and 8b are for 17 October 2014, and Figures 8c and 8d are for 24 October 2014.

Figure 9. (a and c) The values of IG and RZ required to give good agreement between measured sTEC and IRI-derived sTEC for the GPS receiver located at Hailsham, UK. (b and d) The foF2 values obtained from the IRI with the sTEC-derived IG and RZ values compared with the measured foF2 values at Fairford, UK. Figures 9a and 9b are for 17 October 2014, and Figures 9c and 9d are for 24 October 2014.
Developments are required to provide an automated method suitable for use in our application. Once the background parameters have been determined from the TEC measurements, deviations from the background levels may then be employed to estimate prevailing patch parameters, including their size, location, and intensity.

4. D Region Absorption

Empirical models of HF absorption have been developed for the auroral regions [e.g., Foppiano and Bradley, 1985] and for the polar ionosphere during solar proton events [e.g., Sauer and Wilkinson, 2008]. Polar cap absorption (PCA) events may produce several decibels of cosmic noise absorption, measured by riometers at 30 MHz, that can persist for several days [Bailey, 1964]. Over recent years, the number of riometers in the high-latitude region has increased considerably, with 23 stations now operational in the Canadian region alone [Danskin et al., 2008], many of which are fitted with the capability of supplying near-real-time (<15 min latency) measurements online. Consequently, new data-assimilative models have been developed in which the parameters of the PCA models are optimized in real time using a weighted nonlinear regression to riometer measurements, and these will be included in the propagation predictions. Details are not included here, and the reader is referred to Rogers and Honary [2015] and Rogers et al. [2015].

5. Concluding Remarks

A ray tracing method has successfully been used to model the effects of the polar ionosphere on HF signals for historical scenarios. In order to enable this model to be applied in nowcasting and forecasting for operational systems (e.g., the prediction of communications with commercial aircraft prior to dispatch (forecasting) and for frequency management during flight (nowcasting)) over a period of a few hours, we are currently incorporating data from a number of sources including ionosondes and GPS to provide real-time estimates of the background ionosphere and the number and intensity of patches.

We have successfully moved to an IRI-based ionospheric model. Coverage plots for the same time as those presented in Figure 2 using the current version of the new model are given in Figure 11. Good agreement is obtained between the two for 8.0 MHz, but there are differences that are particularly noticeable at 11.1 MHz.

Figure 10. (a and c) Histograms of the percentage of cases where the $f_{\text{F}_2}$ values were improved by using the sTEC-derived $\text{IG}$ and $\text{RZ}$ as input to the IRI and where the values remained unchanged to within ±0.2 MHz as a function of local time. The data were binned into 3 h periods centered on the indicated time. (b and d) The RMS errors of the $f_{\text{F}_2}$ values using the monthly IRI values and the values obtained from the IRI with the sTEC-derived $\text{IG}$ and $\text{RZ}$ values. Figures 10a and 10b are for Alpena, Canada, and Figures 10c and 10d are for Fairford, UK. Data are from 1 to 31 October 2014.
Further developments are required in order to improve performance, and further testing is required to compare modeled values with measured values, in particular noting that small changes in the background electron density can lead to significant changes in the coverage, particularly at distances close to the skip distance. It should also be noted that using the IRI with the monthly IG and RZ values gives us a fall-back position should the required real-time data become unavailable at times.

While ionosonde and TEC measurements will help establish the presence and intensity of patches, it is more difficult to determine the exact number of patches and their physical extent (since time and space are convolved in the GPS measurements because both the points where the path from the satellite to the ground intersects the ionosphere, the pierce point, and the patch are moving). However, there will be occasions when the same patch will affect the TEC on several GPS satellites (i.e., where the patch is larger than the separation of the pierce points), which will allow the size of that patch to be estimated. In view of the uncertainty in patch numbers, positions, and intensities, the simulation may be run with a range of different values of the patch parameters in order to establish an ensemble average for the coverage maps.

The predictions are intended for relatively small solar disturbances where being able to use higher frequencies supported by off-great circle propagation, and thereby avoiding the higher absorption at lower frequencies will provide a communication path. During periods of intense solar activity (e.g., where a large coronal mass ejection impacts on Earth), complete absorption of HF signals is expected, and hence, at these times it is unlikely that successful communication in the polar cap will be possible. We are currently investigating the possibility of exploiting sporadic E propagation at times of high absorption. Ritchie and Honary [2009], for example, reported an increase in the median value of the E region critical frequency in the hours after a sudden storm commencement. Although the presence of sporadic E layers can lead to blanketing (i.e., where reflection from the F region is no longer possible), higher critical frequencies allow higher operational frequencies to be adopted and this will lead to a reduction in the absorption.

Figure 11. Predicted signal coverage at 8.0 and 11.1 MHz for a transmitter located in Qaanaaq for 23:30 UT on 17 October 2014 using the IRI-based model. The colors indicate the received signal strength, while the positions of the transmitter (Qaanaaq) and the receiver (Alert) are marked with Q and A, respectively. (a) 8.0 MHz, background ionosphere. (b) 8.0 MHz, one particular patch scenario. (c) 11.1 MHz, background ionosphere. (d) 11.1 MHz, one particular patch scenario.
Acknowledgments
The authors are grateful to the EPSRC for their support of this research through grants EP/K000787/1 and EP/K000797/1. We are also grateful to the University of Tromsø for the ionograms in Figure 3 (http://geo.phys.uib.no/ionodata/index.html), to the Global Ionospheric Radio Observatory (Rieinisch and Galkin, 2011) for the $f_{ Maps}$ measurements (http://gigo.uml.edu/didbase/scaled.php), and to the Telecommunications/ICT for Development (7/ICT4D) Laboratory of the Abdus Salam International Centre for Theoretical Physics, Trieste, Italy for the calibrated GPS data (http://t-ict4d.licit.it/ nequick2/gps-tec-calibration-online). Our own data may be requested from the corresponding author. This research may be requested from the Telecommunications/ICT for Development (7/ICT4D) Laboratory of the Abdus Salam International Centre for Theoretical Physics, Trieste, Italy for the calibrated GPS data (http://t-ict4d.licit.it/ nequick2/gps-tec-calibration-online). Our own data may be requested from the corresponding author.

References