SAND93--1388
Unlimited Release
Printed August 1993

Distribution
Category UC-~~003~~
705

USING COMPUTERIZED TOMOGRAPHY TO DETERMINE
IONOSPHERIC STRUCTURES

Part I: Motivation and Basic Approaches

Charles N. Vittitoe
Radiation and Electromagnetic Analysis Department
Sandia National Laboratories
Albuquerque, NM 87185

Abstract

Properties of the ionosphere are reviewed along with its correlations with other geophysical phenomena and with applications of ionospheric studies to communication, navigation, and surveillance systems. Computer tomography is identified as a method to determine the detailed, three-dimensional distribution of electron density within the ionosphere. Several tomography methods are described, with a basic approach illustrated by an example. Limitations are identified.

MASTER

REPRODUCTION OF THIS DOCUMENT IS UNLIMITED

Acknowledgments

My acknowledgments and thanks to the following people for their contributions to this project:

Jack Klobuchar, Phillips Laboratory, Jeffrey Austen, Tennessee Technological University, and Ed Fremouw and Jim Secan, Northwest Research Associates (Bellevue, WA), for their helpful discussions in guiding exposure to tomography methods.

Glenn Elliott (SNL 9225). His suggestions led to consultations with Bob Massey, Tom Armstrong, and Tim Murphy, all of Los Alamos National Laboratory, and to examination of an 800-km orbit in attempts to take advantage of the LANL Blackbeard experiment.

Frank Biggs (SNL 9312), Lynn Kissel, Lawrence Livermore National Laboratory, and Tom Marking (formerly with SNL 9312) for helping to increase my efficiency and success with the use of their UFO code for the required unfolding.

Terri Calton (SNL 9352), Dan Carroll (9221), and Mike Sharp (9232) for generating the capability to retrieve GPS W-sensor data from the SNL facilities.

Mel Olman, formerly with SNL's Supercomputing Systems, for leading the battle to resolve several graphics problems.

Norm Blocker (SNL 9206) and Andy Boye (SNL 9225) for their helpful discussions.

This work was supported by Department of Energy, Laboratory Directed Research and Development (LDRD) funds, administered by V. Narayanamurti, a former chairman of the SNL LDRD Oversight Group. Part I of the work was supported by part-time effort during FY91. The research for Part II was concluded by part-time efforts in the following year.

Acronyms

ART	algebraic reconstruction technique
CT	computer tomography
EMP	electromagnetic pulse
ESF	equatorial spread F
GPS	Global Positioning System
HF	high frequency
IES	Ionospheric Effects Symposia
LANL	Los Alamos National Laboratory
MART	multiplicative algebraic reconstruction technique
RF	radiofrequency
SIRT	simultaneous iterative reconstruction technique
SNL	Sandia National Laboratories
SOR	successive overrelaxation
TEC	total electron content
UHF	ultrahigh frequency
UFO	a code for interactive utility programs
VHF	very high frequency
1D, 2D, ...	one dimensional, two dimensional, etc.

Table of Contents

Contents.....	6
Introduction.....	7
What is the ionosphere?.....	7
Why study the ionosphere?.....	9
Contenders for satellite-collected TEC data.....	13
GPS W-sensors.....	13
Blackbeard project sensors.....	14
Is CT reconstruction of the ionosphere feasible?.....	16
Tomography approach.....	17
Some limitations to the approach.....	23
ART and SIRT methods.....	26
ART.....	26
SIRT.....	29
Comparison with MART.....	30
Alternate approach.....	31
Example: Reconstruction from 800-km orbit and one transmitter.....	32
Summary and Conclusions.....	37
References.....	39
 Figures	
1 Two-dimensional tomography geometry.....	20
2 An example of a tomography geometry.....	24
3 Simple electron-density model for 100- to 800-km altitudes.....	33
4 Path and pixel geometry for single-transmitter example.....	34
5 Illustration of paths for a four-transmitter example.....	35

USING COMPUTERIZED TOMOGRAPHY TO DETERMINE IONOSPHERIC STRUCTURES

Part I: Motivation and Basic Approaches

Introduction

The goal is to develop and illustrate procedures to derive the three-dimensional (3D) distribution of the electron density above the contiguous United States by using a computer tomography (CT) analysis of RF transmission data received by satellites. Those data contain information related to the total electron content (TEC in electrons/m², the integral of the electron density over the path from the ground-based pulser). The method is analogous to x-ray tomography used in medical imaging. The report is divided into two parts. Here in Part I we concentrate on motivation and basic procedures. Part II selects an orthogonal set of basis functions to represent the troublesome vertical structure and treats curved paths to increase vertical resolution.

What is the ionosphere?

The ionosphere is produced by the interaction between the earth's first layer of defense (the atmosphere) and the harsh environments propagating from the sun. The fully ionized plasma sea and the photon shower emerging from the sun engulf the neutral atoms and molecules at high altitudes in the earth's atmosphere, stripping away electrons to form the ionosphere. Galactic cosmic rays add to the severity of the environment. It is a dynamic interaction, constrained both by atmospheric dynamics below and by deep space and solar plasma physics above. The earth's magnetic field coupled with the atmosphere gives a second layer of protection from the hot tenuous plasmas of the magnetosphere and the solar wind and links these plasmas to the ionosphere. We refer the reader to Kelley and to the extensively revised Davies for a fascinating introduction to the ionosphere.¹⁻² Some of these basics will be required later.

Although quiet, reproducible profiles of electron density characterize the ionosphere at times, the ionospheric profiles

can become chaotic. A series of regular perturbations also occur, with periods ranging from the 11- and 22-year solar cycles, through less noticeable semiannual and annual effects (that are influenced by the inclination of the earth's axis to the ecliptic plane and variation in the earth-sun distance), through the 27-day solar-rotation period, to just a few seconds for development of some instabilities associated with sunrise or sunset. Especially long and irregular "cycles" are associated with the six reversals in the main geomagnetic field that have occurred in the last million years.³ The most dramatic regular perturbation is the twice-daily transition between nighttime and daytime ionospheres, with added ionospheric distortion caused by the 11° tilt and the offset between the earth's magnetic dipole and the earth's rotation axis. More erratic variations are associated with geomagnetic activity, particle precipitation, and atmospheric disturbances.

Examples of chaotic behavior within the ionosphere include the F-region equatorial spread F (ESF) instabilities.¹ One of these ESF instabilities is the plasma analog of the classical Rayleigh-Taylor instability that occurs when a heavy fluid is supported against gravity by a lighter fluid. A steep upward gradient of electron density develops on the bottomside of the nighttime F layer. The gravitational force is downward and the plasma is nearly collisionless. The plasma fields that build up from a slight perturbation cause the lower density plasma to be buoyed upward toward a region with higher electron density. This creates a larger perturbation and drives the instability farther. The system is unstable any time that gravity and the gradient in electron density are oppositely directed. In addition to gravity, ambient electric fields and neutral winds can also lead to similar instabilities. The resulting upwellings of plasma, at times referred to as auroral blobs or polar cap patches, can develop and rise about 100 km within 10 minutes.⁴ The blobs rise until the electron density within the bubble becomes greater than or equal to the ambient density. The phenomenon occurs mainly at night and infrequently during daylight hours.

Why study the ionosphere?

In 1987, the Naval Research Laboratory, the Office of Naval Research, the Air Force Geophysical Laboratory, the US Army Communications-Electronics Command, the Defense Nuclear Agency, the Naval Ocean Systems Center, the Institute for Telecommunication Sciences, and the Voice of America all cooperated to hold a symposium on the Effects of the Ionosphere on Communication, Navigation, and Surveillance Systems.⁵ This was the fifth in a series of such meetings, generally referred to as Ionospheric Effects Symposia (IES). The fourth symposium covered communication, command, control, and intelligence systems, and selected papers concentrating on radio systems were published in 1985.⁶ These symposia provide a good introduction to ionospheric effects on systems. The papers presented at the Electromagnetic Wave Propagation Panel Symposium summarize military interests.⁷

One of the obvious communication systems affected by perturbations in the ionosphere is a long-range transmission system that communicates by ionospheric reflections. The transmitted power required to establish intelligible reception is more often determined by ionospheric noise than by receiver limitations. Even at higher frequencies such as in the VHF/UHF bands at 30 to 1000 MHz, where reflections do not usually occur, the ionosphere can inhibit radio communication and cause interference.⁸⁻⁹ At the dawn of satellite telemetry, frequencies near 137 MHz were used in attempts to avoid ionospheric effects. Almost immediately, two degradations were found: (1) Faraday rotation giving polarized signals and associated fading, and (2) amplitude fluctuations caused by irregularities.² Some satellite communication systems account for known irregularities in the ionosphere.⁸ Those perturbations that are chaotic or that might be artificially injected (such as HF or microwave injection, nuclear or chemical seeding, or charged particle injection) are harder to process routinely.

The irregularities in the ionospheric electron density cause random phase fluctuations in the radio waves from satellites or in the radio-astronomy wavefronts from galaxies. At the ground, phase mixing and amplitude fluctuations develop. These temporal fluctuations in phase and amplitude are called

scintillations. They are a severe source of noise for VHF/UHF satellite communication links. The associated ionospheric irregularities in electron density are thought to occur naturally in the F region of the ionosphere, with the dominant fluctuations at the magnetic equator shortly after sunset, at the nighttime auroral oval, at the dayside cusp, and at the region within the polar cap.⁹ In the 1960s, communication bands were shifted to 4-6 GHz to avoid these scintillations.² Although the distortions were reduced, they were still troublesome in auroral zones and in the equatorial region at certain times and seasons.

Without irregularities and fluctuations, the TEC along the propagation path still alters the signal transmission. A Doppler radar suffers a Doppler frequency shift that depends on the time rate of change of the TEC along the intervening path as well as on the radial velocity of the target.¹⁰ If the electron-density profile is known, a correction can be made for the ionospheric time delay, and even for the known portion of phase scintillation caused by the ionosphere. Such corrections are needed for certain radar applications such as for satellite detection radars.¹¹ In the system-design stage of satellite development, a global model of TEC is often used to estimate time delays, refraction effects and Faraday rotation effects, along with expected variations as the ionosphere changes.² (Some authors restrict TEC to refer to the total electron content along a vertical path. Then slant TEC refers to the TEC along a more general, nonvertical path. We use TEC to indicate the integral of electron density along a general path and use vertical TEC when the path is restricted to be vertical.)

The Global Positioning System (GPS) satellites are being augmented by Block IIA satellites equipped with W sensors to detect electromagnetic pulses that help locate nuclear bursts in the atmosphere. These sensors use a set of look-up tables to correct the signal arrival for delays caused by the TEC along the propagation path and by the electronic circuitry.¹²⁻¹³ To first order this is a straightforward procedure. However, higher-order terms alter the association between the delays and the TEC, so the look-up tables might need updating and may not be able to account for TEC alterations that occur on a time scale shorter than the look-up table recycle time. Changes resulting from solar flares or from equatorial shift F phenomena

might introduce errors in the source-satellite distance. Thus, a further impetus for studying the ionosphere is to examine the magnitude of these higher-order effects, test for consistency between the measured ionosphere and the look-up tables, and introduce any needed compensation or increase in uncertainties required by the effects.

Recent discoveries indicate that global averages of several atmospheric parameters (such as CO, CO₂, O₃ densities) are changing with time. Added data are required to identify sources of these changes. We expect that some of the possibilities are correlated with changes in the ionosphere. If global snapshots of the ionospheric structure were obtained in a cost-effective manner, then the ionospheric study would be justified and the data gained might reveal correlations that help identify basic drivers for the changes.

As an example of a correlation between the ionosphere and other geophysical phenomena, consider the weather patterns. At first glance the two phenomena appear independent. The troposphere continues up to about 10-km altitude, where the temperature starts to rise with altitude as ozone filters out the ultraviolet rays from the sun. The stratosphere continues from 10 km to about 50 km at the stratopause, where radiative cooling causes the temperature to again start decreasing with an increase in altitude. The D region of the ionosphere is just starting at about 50 km, so the ionosphere is above the stratosphere. The stable layers at the tropopause and at the temperature minimum (the mesopause) near 80 km reduce mixing between the layers. Our weather occurs in the troposphere, below 10-km altitude. Near the top of the troposphere we find the (50 to 100 m/s) jet streams with their strong correlation with weather.

A few triggers for formation of ionospheric layers are still being debated.¹⁴ Some believe that x-rays from solar flares, gamma-ray and x-ray bursts, as well as supernova explosions, have given measurable D-region perturbations.¹⁵ Evidence is strong for correlation between weather features and sporadic-E formation (occasional enhancements in electron density at 100- to 120-km) well above the tropopause.[Kelley,¹ p 224] Sporadic-E layers may be triggered by jet streams, thunderstorms, weather fronts, tornadoes, ridge patterns in the

upper atmosphere, and volcanic eruptions. Even blast waves from large manmade chemical explosions have induced oscillations in the ionospheric E layer.¹⁶ A 1990 study of data for the winter of 1983/1984 concluded that meteorological factors govern the behavior of the lower ionosphere in Europe.¹⁷ Solar and geomagnetic activity contributed only weak influences.

A disturbance generated below the tropopause can spread upward through the tens of kilometers to the E region. The vertical stability at the tropopause does not inhibit horizontal motion. The stability generates restoring forces that form waves in the flow of air above some disturbing obstacle. These waves near the top of the jet stream drive shear waves at higher altitudes. Conservation of energy within the perturbation then causes an increase in perturbation speed as the air density decreases with height.[Kelley,¹ p 226] The ionospheric plasma (carried by the winds at the shear) reacts with the earth's magnetic field and moves the charges vertically. The wind-shear theory initially was questioned because recombination rates were too high to support electron densities near 10^6 electrons/cm³ that were sometimes experienced. Later work indicates that Mg⁺, Si⁺, and Fe⁺ caused by meteor showers are the dominant ions driving the long-lived sporadic-E layers.[Kelley,¹ p 233]

The ionosphere is also coupled to geophysical phenomena at higher altitudes. Above the mesopause, near 80-km altitude, the temperature increases to values often well above 1000 K. This is the thermosphere that interactively couples to the ionosphere. Accurate electron densities are required to estimate conductivity and joule heating effects. The ionospheric plasma distributions are affected by the neutral winds and the electric fields generated by typical dynamo action of charged particles moving in magnetic fields. The neutral winds also respond to the drag force imposed by collisions with ions constrained by the geomagnetic field. The structure is described by coupled dynamic, thermodynamic, chemical, electrodynamic, and continuity equations for the neutral and ionized components of the upper atmosphere above 80 km. A good determination of the variation in plasma density within the ionosphere would provide realistic constraints on attempts at solving these coupled equations.

Contenders for satellite-collected TEC data

GPS W sensors

Block IIA satellites are now joining the GPS constellation. Each is equipped with a set of VHF receivers, called a W sensor, designed to locate nuclear bursts in the atmosphere by detecting EMP signal arrival in three 2-MHz-wide frequency bands in the interval from 100 to 150 MHz. The bands are separated by 10.23 MHz. About two dozen satellites are planned for the constellation, each flying at equal spacings around six orbital planes and at about half the earth-satellite distance for a geosynchronous orbit. The sensors record the difference between arrival times for the three frequency bands. These data allow evaluation of the TEC over the path between the source and the sensor. The resultant delay times then identify the geometric path length. By combining path lengths from at least four satellites, the EMP source position and detonation time are determined.

The Air Force Systems Command has built an RF pulser to calibrate the W sensors. This calibration refines the set of update tables stored within the satellite electronics. The update tables relate delay times to specific arrival-time corrections and to the resulting TEC. **A few such calibrating RF pulsers could provide sources for a three-dimensional reconstruction of the ionosphere, using a method analogous to computer tomography (CT).** The TEC data from a series of sources and detected by the few dozen GPS receivers as they orbit the earth provide just the type of path-integral data needed for CT reconstruction of the ionosphere. Even with only a few GPS satellites equipped with W sensors deployed within the next few years, assumptions about ionospheric shape should allow some reconstruction with just the one calibrating pulser planned for Florida and the one pulser built by Los Alamos National Laboratory in New Mexico. TEC data may also be available from phase detectors on GPS satellites that work in an interferometer mode to detect phase shift on L1 (1.6 GHz) and L2 (1.2 GHz) signals to ground-based receivers.

One limitation to using GPS data for CT reconstruction is the slow orbital speed of satellites at one-half geosynchronous altitude. At a distance r from the earth's center, the angular

frequency about the orbit is proportional to $r^{-3/2}$. When r is the geosynchronous distance, this frequency is one revolution per day. At one-half the geosynchronous distance, this frequency is $8^{1/2}$ revolutions per day, or $42.4^\circ/\text{hr}$. The earth's rotation rate is $15^\circ/\text{hr}$. So, in the several hours required for a horizon-to-horizon pass of a GPS satellite, the ionosphere has considerable time to change.

A solar storm can inject about 4×10^{17} electrons/ m^2 into a vertical TEC.¹⁰ Such a perturbation can rise to its peak value in one hour and can remain for about six hours.¹⁰ The ESF phenomena can elevate a blob of plasma 150-km thick with $\sim 10^5$ electrons/ cm^3 upward for ~ 200 km, into a region with ambient densities a few times larger. Such motion can occur within hours and can alter vertical TEC values by $\sim 1.5 \times 10^{16}$ electrons/ m^2 . CT methods based on slow-moving GPS satellites cannot hope to time-resolve such fine detail.

One method to get finer detail was mentioned by John Klobuchar of the Phillips Laboratory, Hanscom AFB, Massachusetts. A transmitter (or the right receiver) could be placed on a satellite in circular orbit at a few hundred kilometers altitude. By transmitting to the GPS satellites, CT data to probe altitudes above the transmitter could be generated much faster. Klobuchar reported that Air Force agencies have received such proposals, but that the modeling work was not done. Some progress was reported at the 1991 North American Radioscience Meeting.¹⁸ Such a system would be quite expensive. Although CT data from the GPS constellation and calibrating pulsers in New Mexico and Florida would limit study of the ionospheric structure to longer time averages, the cost is smaller.

Blackbeard project sensors

The Blackbeard project at Los Alamos National Laboratory (LANL) has features that make it a contender to generate TEC data for ionospheric reconstruction. Discussions with Tom Armstrong and Tim Murphy at LANL indicate that the satellite will be launched into a 741-km orbit and will carry receivers to detect 35-MHz signals within narrow bands. The orbit inclination to the equatorial plane will likely be between 50°

and 65° (depending on compromises with the main satellite missions). The satellite speed in orbit is $2\pi(741+6371)$ km/5976.3 s, or 7.48 km/s. This is $217^\circ/\text{hr}$. Such a speed allows probing with better time resolution than available from GPS satellites.

Four transmitters generate 35-MHz radiation with 100-Hz separation between their emissions. The transmitters are deployed on the ground, in a square mesh with 200-m sides. An elliptical polarization is employed so that only the ordinary rays (in the birefringent ionosphere) are detected. Doppler shifts caused by the satellite motion are also corrected. (This requires the receivers to have a wider bandwidth to account for the shifts.) For our purposes, the net result is four continuous TEC measurements for paths between the four transmitters and the satellite as it moves overhead. The accuracy goal is TEC within 1 percent. The measurement can be made for satellite zenith angles up to about 60° ; that is for about 320 s of an overhead pass (2400-km long). As paths pass by 400-km altitudes (near the expected peak electron densities) they span a 1400-km horizontal distance.

The deployment of the transmitters within a kilometer of each other is not optimum for CT reconstruction. Separation along a 200-km baseline would allow better resolution of the 10- to 200-km wavelengths that might appear in the ionospheric structure. These might be called wide-band distortions in the ionosphere. Discussions with Tom Armstrong at LANL suggest that structures with 30-km wavelengths occur often, whereas 10-km wavelengths are thought useful to detect. Perhaps if CT gives good resolution of 10-km wavelengths (requiring about 1-km resolution of the wavelength structure) greater resolution would be sought. Interferometer types of measurements (that account for the narrow spacing of the transmitters) are planned by LANL to detect the more narrow-band distortions in the ionosphere.

In addition to moving the transmitters to increase CT resolution, another obvious addition would be a sounder designed to work with the same transmitters to record TEC data from the transmitters to positions along a vertical path at the edges of the tomography plane being probed. Such advantages will become apparent later.

The tomography mesh suggested by this Blackbeard satellite geometry is about 200 km (to detect the 200-km wavelengths horizontally) by about 300 km (to detect the main structure in the ionosphere from 300- to 600-km altitude). With just 40 bins horizontally and 60 bins vertically, the mesh resolution would be near 5 km on a side. This is sufficient to resolve 50-km wavelength structures and even to suggest structures down to 25-km wavelengths. Shorter wavelengths will require further refinements, shorter base lines, and/or more accurate TEC measurements. The mesh also requires new deployment of the transmitters along the 200-km baseline. The 35-MHz frequency requires further analysis to include curved paths in the tomographic reconstruction.

Is CT reconstruction of the ionosphere feasible?

In 1988, J. R. Austen and co-workers at the University of Illinois demonstrated the feasibility of using CT techniques for reconstructing an image of the ionosphere.¹⁹ Two years later this same group, aided by data from J. A. Klobuchar and J. Stalker of the Phillips Laboratory, successfully applied the technique to simulated TEC data for a section of the ionosphere from 50- to 1000-km high and over 16 degrees of latitude.²⁰ A further extension by the same group included some effects of expected noise in the data.²¹ These initial attempts were for two-dimensional images in a latitude-altitude plane. They are the only published data found on CT reconstruction of the ionosphere. The extension to three dimensions is straightforward--just extend the two-dimensional probing to neighboring planes. (There are three-dimensional techniques under development, but these are on the research forefront - Private communication, R.M. Leahy, University of Southern California, October 1990.)

In these reconstructions¹⁹⁻²¹ the satellite at 1000-km altitude transmitted downward toward three to five receivers on the earth's surface. Straight integration paths were assumed. The GPS geometry has receivers in satellites at very high altitudes with the few transmitters on the ground. At 1000-km altitude, the orbital period is 1.75 hours. So traversing 16 degrees of latitude (in a polar orbit) requires only 4.6 minutes. The ionosphere is considered stationary over such a

short time. The GPS satellites have an 8.5-hour period--too long to expect a stationary ionosphere. Again, when GPS data are used, we must accept the limitation of determining average structure with the ground-based transmitters. The feasibility of reconstruction might also be questioned for this broader GPS geometry. In the following section, we review the CT methods used and examine alterations for the GPS geometry.

The published reconstructions¹⁹⁻²¹ were limited to altitudes below 1000 km. This region is the most active portion of the ionosphere, where the D, E, and F regions reside. Above 1000 km, the plasma density is below 10^{10} electrons/m³, with an e-folding scale height of 200 km or larger. Below 60 km, the density is less than 10^9 electrons/m³. The peak densities (above 10^{12} electrons/m³) occur in the F region. (With 10^{10} electrons/m³ at 1000 km and with an e-folding scale height of 200 km, it would take 4.6 scale heights or 1920 km altitude to reach the interplanetary density of 10^8 electrons/m³. The associated vertical TEC from 1000 to 1920 km is 2×10^{15} electrons/m².) Thus, most of the contribution to 10^{17} - 10^{18} electrons/m² TECs through the ionosphere comes from altitudes between 100 and 1000 km.

At the 1991 meeting of the International Union of Radio Science, groups from Russia and Wales reported on successful reconstructions using experimental data from the Navy Navigation Satellite System.²²⁻²³ Hence, not only is CT reconstruction of the ionosphere feasible, it has been accomplished!

Tomography approach

The basic problem in CT is to reconstruct a function using data for a series of line integrals through the function. As an example, we consider a two-dimensional slice through the ionosphere. The electron number density is represented by $n_e(x,y)$ and is a function of the two coordinates x and y . A line integral along path p through the $n_e(x,y)$ gives the total electron content (electrons/m²) along that path, T_p .

$$T_p = \int_p n_e(x,y) ds,$$

where ds is the element of path length for the propagation path p in the x,y plane. In this case the T_p values are referred to as the projection data. Knowing T_p for many p gives the basic projection data needed by CT to reconstruct $n_e(x,y)$.

One complication to computer tomography applied to the ionosphere is that the paths p are usually curved. The curved paths are common to several other CT applications, such as with seismic waves or ultrasound techniques. The W sensor detects in the 100- to 150-MHz band where curvature effects are small. At narrow bands near 35 MHz, used in the Blackbeard project, the effects can be large.

Let us apply a perturbation technique by constructing a model of the ionosphere, such that

$$n_e(x,y) = n_o(x,y) + f(x,y),$$

where $n_o(x,y)$ is an initial guess and $f(x,y)$ is the sought perturbation.

$$T_p = \int_p n_o(x,y) ds + \int_p f(x,y) ds.$$

The T_p data can then be modified by subtracting the known portion.

$$T'_p = T_p - \int_{p_o} n_o(x,y) ds = \int_p f(x,y) ds + \int_{p-p_o} n_o(x,y) ds.$$

With $n_o(x,y)$ known, the corresponding path p_o can be determined and can be used as an approximation to p . If $f(x,y)$ is sufficiently small, the difference between path p and p_o can be neglected, and the $f(x,y)$ can be determined from usual CT techniques. If not, some iteration with the model $n_o(x,y)$ will be required for convergence. If $n_o(x,y)$ is a constant, the paths will be straight lines and we return to a more usual CT problem.

In addition to being able to account for the curved path by using a model ionosphere, another advantage of using a model is to help unfolding when the projection data are incomplete. With only a few radiation sources, the line integrals are not

available for all angles through the x,y plane. The paths that are available may not adequately sample the entire x,y plane where $f(x,y)$ is significant. Such shortcomings give an incomplete set of projection data. Incomplete data problems are the rule, not the exception in CT.

An added advantage of using a model $n_0(x,y)$ for the ionosphere is that the model helps limit the altitude over which the $f(x,y)$ is sought. The interplanetary electron density is expected to be near 100 electrons/cm³. At satellite-orbit radii near 2×10^7 m, such a background could perturb the T_p by a few times 10^{15} electrons/m². The model allows the background electrons to be easily included in the analysis. The $f(x,y)$ sought is then confined to altitudes much less than 2×10^7 m (for a case with the satellite at 2×10^7 m).

When the reconstruction region is between the altitudes 50 and 1000 km, the lower density limit is near 10^9 electrons/m³. With densities higher than 10^{12} electrons/m³ also present, the n_0 values span a wide range. A normalized TEC could be formed as

$$T_p = \int_p n_e(x,y)/10^9 \, ds,$$

to reduce the variability in the sought $n_e(x,y)$ to values between 1 and 1000 rather than between 10^9 and 10^{12} . However, with such a change of units the range of values is still large and no advantage is apparent. Another way to reduce the range is to replace the normalized electron density by its logarithm. But the analysis would then require a measurement (analogous to T_{np}) that depended on the line integral of the logarithm of the normalized density. Such a measurement is not readily available. The problem appears destined to require unfolding $n_e(x,y)$ over a large dynamic range.

A simplified geometry is illustrated in Figure 1. Signals are transmitted from n pulsers located on the earth's surface and are collected at m equally spaced receiver positions along a satellite orbit. The ionosphere to be reconstructed is confined to a region between altitudes h_1 and h_2 and between latitudes L_1 and L_2 . Here the orbit is over the poles and the pulser is assumed to be in the same vertical plane as the orbit. The h,L

coordinates in the CT plane have been converted to x, y Cartesian coordinates, with the origin at 65° N latitude at ground level and with y vertical. The region being probed is divided into n_h horizontal strips and n_L vertical strips, illustrated in Figure 1 for $n_h=4$ and $n_L=6$. Intersections between these finite-width strips form a set of $n_h \cdot n_L = N$ pixels, with each pixel assigned a value of electron density. One basic assumption is that $n_e(h, L)$ is effectively constant inside each pixel. There are $n_L \cdot n_h$ unknown values of $f(h, L)$ to be determined. In Figure 1 the f_i are arranged row by row so that the $f_1, f_2, \dots, f_{n_L}$ correspond to the first row of the section being probed; $f_{n_L+1}, \dots, f_{2n_L}$ corresponds to the second row, etc. The subscripts on the f_i are printed on the pixels in Figure 1 to identify the pixel number for each pixel. By detecting m signals from each of n transmitters on the ground, we generate $n \cdot m$ relations connecting the $n_L \cdot n_h$ unknowns.

Figure 1. Two-dimensional tomography geometry.

In Figure 1, note that with the one transmitter indicated, with the six receivers, and with the paths indicated, at some locations the perturbations to the model ionosphere are not probed. Estimates for f_{12} , f_{17} , f_{18} , f_{22} , f_{23} , and f_{24} cannot be improved at all. Since the model is the initial guess and no paths cross these pixels, no pertinent data are generated and these pixel values remain zero. Pixels 6, 11, 16, and 21 suffer from being intercepted by only a single path, hence their pixel estimates cannot be improved relative to each other. Despite these limitations to a few pixel pairs relative to each other and to a few pixels that must remain zero, the basic T_p data from just one receiver will allow iteration among the pixel values, f_i , for i ranging from 1 to N , to improve the estimates of electron density. By pulsing more frequently, the number of receiver positions will increase, increasing the resolution and/or reducing the number of limitations. If the T_p' data are zero for each path p , then the model $n_o(h,L)$ is completely consistent with the projection data and more information is needed to improve estimates from the model itself.

The line integral associated with path $p_{j,k}$ connecting transmitter j ($1 \leq j \leq n$) with the receiver k ($1 \leq k \leq m$) has the form

$$\int_{p_{j,k}} f(h,L) ds = \sum_{i=1}^N f_i r_i(j,k) = T'_{j,k}.$$

For a fixed j,k the $r_i(j,k)$ represent the length of the path $p_{j,k}$ as it intersects the pixel associated with f_i . If no intersection occurs for a given pixel, the corresponding $r_i(j,k)$ is zero. The quantities $T_{j,k}'$ are the data being processed, that is, the TEC for path j,k minus the line integral of the model electron density, integrated over the path j,k (for these straight-line paths).

The set of $m \cdot n$ equations for each of the paths can be represented as the matrix equation:

$$R f = T,$$

where

$$R = \begin{bmatrix} r_{1,1} & r_{1,2} & \dots & r_{1,N} \\ \vdots & \vdots & & \vdots \\ r_{mn,1} & r_{mn,2} & \dots & r_{mn,N} \end{bmatrix}.$$

As indicated in R , when the subscripts in $r_{i,j}$ can be confusing, the subscripts are separated by a comma to aid interpretation. T is a column matrix, with elements equal to the $T_{j,k'}$ values for the $m \cdot n$ paths. R is a $m \cdot n \times n_h \cdot n_L$ ($=N$) rectangular matrix, referred to as the tomography matrix. The elements of a row of R are the N values of $r(j,k)$ (the path length within each pixel) that correspond to the path j,k . Subscripts on the elements of R represent the path (row) and the pixel element (column). We are trying to find the column vector f , having N elements that give the magnitude of the perturbation to the electron density in each pixel.

We follow the ordering scheme discussed by Koltracht, Lancaster, and Smith.²⁴ The $m \cdot n$ rows of R are ordered such that data for source number 1 corresponds to the first m rows, row i (for i varying from 1 to m) being for the path from source 1 to receiver position i . If there is a second transmitter, the next set of m rows of R corresponds to source 2 transmitting to the same set of m receivers. Subsequent rows use the same ordering scheme for each additional transmitter. Thus, the k th row of R corresponds to the source-receiver pair (i,j) where $k = (i-1)m + j$. As an example, we list (in km) the nonzero elements of the tomography matrix corresponding to Figure 1 (where only one source is present):

```

r1,1 =225.0 r1,7 =225.0 r1,13=225.0 r1,19=225.0 r2,1 = 91.0
r2,2 =148.2 r2,7 =240.3 r2,13=241.4 r2,19=242.7 r3,2 = 99.2
r3,3 =176.2 r3,8 =280.1 r3,13=123.2 r3,14=162.2 r3,19=291.7
r4,3 =104.9 r4,4 =216.6 r4,9 =333.0 r4,14=340.3 r4,15= 7.1
r4,19=296.0 r4,20= 69.5 r5,4 =102.1 r5,5 =265.4 r5,9 = 87.1
r5,10=302.3 r5,14=126.7 r5,15=292.1 r5,19=229.6 r5,20=231.1
r6,5 = 87.3 r6,6 =320.6 r6,10=145.9 r6,11=296.6 r6,15=275.3
r6,16=217.9 r6,19=173.5 r6,20=332.1 r6,21= 7.1

```

Because paths do not intersect the pixels, columns 12, 17, 18, 22, 23, and 24 are all null columns in the tomography matrix. If three more transmitters were added to Figure 1, the

tomography matrix would then be square, with 24 paths (rows) and 24 pixels (columns).

The tomography matrix, R , is large by most standards. The number of pixels $n_H \cdot n_L$ is expected to be greater than 100. The number of transmitters times the number of receiver positions can easily give more than 100 paths. In Austen's early work on ionospheric tomography, he choose $N=120$ pixels and $m \cdot n=165$ paths.¹⁹

The tomography matrix is a sparse matrix. Although each row of R has length $n_H \cdot n_L$, there are no more than $n_H + n_L - 1$ nonzero entries in each row. For random path orientation, we expect about $(n_H \cdot n_L)^{1/2}$ nonzero elements in a row. In Austen's row length of 120 elements, only about 11 elements were nonzero.¹⁹ In Figure 1 the row length is 24; at most 7 row elements are nonzero.

Some limitations to the approach

The matrix rank is the number of linearly independent row vectors that constitute the matrix. The tomography matrix is not expected to have full rank. That is, the rank is expected to be less than $m \cdot n$. In fact, a row with all zeros is quite possible, especially with a poor choice of path geometry. Even without a null row, the source and receiver positions can be such that the sum of two rows of R is just a scalar multiple of the sum of two other rows. Such dependencies or null rows prevent $Rf=T$ from having a unique solution f , as illustrated later. As a result of errors in TEC or in modeling, the system $Rf=T$ may even become inconsistent.

A minor of a rectangular matrix is the determinant obtained from an equal number of rows and columns of the matrix. If a matrix has rank r , then there must be at least one nonzero minor of order r and no nonzero minor of order greater than r . Rounding errors are likely to confuse any numerical method to determine the rank of a matrix.

Koltracht, Lancaster, and Smith²⁴ give a good example of a tomographic reconstruction where the sum of two rows of R is a scalar multiple of the sum of two other rows, making the rank of

R less than $m \cdot n$. We examine their example with minor alterations to be consistent with our notation.

With unit-square pixels in an h, L plane, and with two transmitters (S_1, S_2), two receivers (R_1, R_2), and a 2×2 array of pixels, the geometry is illustrated in Figure 2. The tomography matrix for this configuration is

$$R = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & \alpha & \alpha & 0 \\ \alpha & 0 & 0 & \alpha \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

where $\alpha = (5/4)^{1/2}$. Here, row 1 plus row 4 is a scalar times the sum of rows 2 and 3. The rank of this R is less than 4. Even though the matrix equation $Rf=T$ involves four equations in the unknown parameters f_i , $i=1,2,3,4$, a unique solution is not possible because the four equations are not independent. R is not a positive definite, square, symmetric matrix with an easily identified inverse that would make the solution to $Rf=T$ much more straightforward. In fact, R is a singular matrix.

Figure 2. An example of a tomography geometry.

One minor of R is the determinant of a matrix having the rows $(1 \ 0 \ 1)$, $(0 \ \alpha \ \alpha)$, and $(\alpha \ 0 \ 0)$. Since this minor is $-\alpha^2$ and is not zero, the rank of R is 3 in our simple example. Our matrix equation $Rf=T$ then has one nontrivial solution ($m \cdot n$ -

rank) for which $Rf=0$. Because of the simple structure, the f is easily identified by

$$R \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

The solution to the homogeneous problem $Rf=0$ is referred to as the kernel of R , denoted by

$$\text{Ker}R = \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}.$$

Any scalar multiple of $\text{Ker}R$ is a solution to the homogeneous problem and could be added to transform from f to $f + \text{Ker}R$, while still being consistent with the tomography data. If $n_e(x,y)$ were treated directly rather than $f(x,y)$, the elements in the column matrix would be restricted to positive values and this $\text{Ker}R$ could not inadvertently add to our solution. Since we are solving for the column matrix f , both positive and negative elements are allowed. Precautions are needed to prevent the elements of f from being contaminated by $\text{Ker}R$ contributions during the solution procedure. Such pollutions are called "ghost images."²⁴

The electron-density variation in the ionosphere is more dramatic vertically (with its layered structure) than horizontally. If we ignore the horizontal structure in Figure 1, we would have four layers vertically ($n_h=4$) and only one horizontal cell in each layer ($n_L=1$). In that case the tomography matrix with its six rows (paths) and four columns (pixels) calculated from that given for Figure 1 is

$$R = \begin{pmatrix} 225.0 & 225.0 & 225.0 & 225.0 \\ 239.2 & 240.3 & 241.4 & 242.7 \\ 275.4 & 280.1 & 285.4 & 291.7 \\ 321.5 & 333.0 & 347.4 & 365.5 \\ 367.5 & 389.4 & 418.8 & 460.7 \\ 407.9 & 442.5 & 493.2 & 512.7 \end{pmatrix}.$$

If the single transmitter has been located at the center of the distribution ($L=65^\circ$), paths 1, 2, and 3 would give data identical to that from paths 6, 5, and 4. We would then have three paths for the four unknown electron densities. Taking only the first three paths (since the other three paths would then be dependent) and dividing all the elements by 225.0, the tomography matrix is

$$R' = \begin{bmatrix} 1.000 & 1.000 & 1.000 & 1.000 \\ 1.063 & 1.068 & 1.073 & 1.079 \\ 1.224 & 1.245 & 1.268 & 1.296 \end{bmatrix} .$$

The determinant of the first three columns is 0, so the rank is 2 and two of the three paths are (weakly) independent. Note that rounding to two significant figures in the r_{ij} nearly removes this independence for any 2×2 array in the matrix. If the pixels were rectangular in the mesh of Figure 1, that would remove the independence. As the mesh limits L_1 and L_2 come closer together and the pixels come closer to a rectangular shape, there is a tendency toward more dependence. This is a fundamental obstacle.

ART and SIRT Methods

At the University of Illinois, the early efforts to reconstruct the ionosphere used the algebraic reconstruction technique (ART) and the simultaneous iterative reconstruction technique (SIRT).¹⁹ In addition to giving arguments for these choices, Reference 19 also gives references for more general reviews of CT methods. Readers interested in a broader, more detailed explanation of CT methods will want to refer to these more extensive reviews. We discuss the ART and SIRT methods and compare them with a later approach implemented at the University of Illinois.²⁰ We also discuss an alternate approach.

ART

The ART iteration scheme updates the unknown vector f for each component in such a way that the following single equation is satisfied exactly:

$$T_i = r_{ij} f_j , \quad i=1, 2, \dots, N_{\text{path}}.$$

We use a summation convention where repeated indices are summed over all possible values. Here the summation over the N values of j is understood. When it is not so clear which index is summed, the summation will be written explicitly. The next few paragraphs show how the iteration is done.

Let the i^{th} row of R be denoted by the column vector:

$$R_i = (r_{i1} \quad r_{i2} \quad r_{i3} \quad \dots \quad r_{iN})^T,$$

where the superscript T indicates the transpose of the vector. Then the element of projection data can be written as the inner product of the two column vectors, R_i and f .

$$T_i = \langle R_i^T f \rangle,$$

where the elements of R are

$$r_{ij} = \int_{p_{ij}} ds.$$

The i is the path index, whereas j varies over the number of pixels (as the summation over j is performed in the inner product). As before, the path integral represents the path length of path i through pixel j . The r_{ij} is zero if the path does not intersect the pixel; thus, $r_{ij} \geq 0$.

The initial guess for f_j is zero for j varying from 1 to N . At iteration k , we compute the difference between the projection data element T_i and its k^{th} estimate, T_i^k , where T_i^k is found from the current estimate of the solution, f^k .

$$s_i^k = T_i - T_i^k = T_i - \langle R_i^T f^k \rangle.$$

This difference is called the element i in the residual column matrix, s^k . The i index ranges over the $m \cdot n$ set of paths (over the number of rows in R). The k starts with the initial $k=0$, and increases by one for each subsequent iteration. At the next iteration we take

$$f^{k+1} = f^k + \frac{\omega s_i^k}{N \sum_{j=1}^2 r_{ij}^2} R_i .$$

The ω is the relaxation parameter, which is usually in the range of 0 to 2. If we operate on this equation with R_i^T from the left, we find:

$$\langle R_i^T f^{k+1} \rangle = \langle R_i^T f^k \rangle + \frac{\omega s_i^k}{N \sum_{j=1}^2 r_{ij}^2} \langle R_i^T R_i \rangle .$$

Since $\langle R_i^T R_i \rangle = \sum_{j=1}^N r_{ij}^2$, for $\omega=1$ we find the identity:

$$\langle R_i^T f^{k+1} \rangle = T_i = r_{ij} f_j^{k+1} .$$

The iteration scheme for f^k stepping to f^{k+1} satisfies $T_i = \langle R_i^T f^{k+1} \rangle$ exactly when $\omega=1$. The choice of ω affects stability and speed of convergence. The method is directly analogous to the successive overrelaxation (SOR) method of numerical analysis.²⁵ As with SOR, in some applications the ω varies with the iteration index k .

In the iteration procedure, each increase in k is accompanied by an increment in the row of R . The row index is $\text{mod}(k, N) + 1$. We use the standard divide for remainder, where $\text{mod}(k, N) = k - N \cdot \text{Int}(k/N)$. The Int function truncates to the integer value. This forces the i values to cycle through the N pixel values as k increases, starting with iteration $k=0$.

ART has the advantage of versatility. It can be applied to a wide range of geometries and even to incomplete data problems. If the R and T are consistent, then ART converges for $0 < \omega < 2$.²⁵

SIRT

Some applications must use data where errors in T_i and approximations in the evaluation of r_{ij} have caused $Rf=T$ not to be a consistent set of equations. Then convergence cannot be expected. A finer choice in the gridding and an increase in the number of paths might lead to widely differing results rather than to increased resolution. The SIRT methods are designed to give convergence in these more troublesome conditions.²⁶

A general class of SIRT methods is given by²⁶

$$f_j^{k+1} = f_j^k + \frac{\omega}{\gamma_j} \sum_i \frac{r_{ij} s_i^k}{\rho_i}.$$

With the convention that $0^0=0$ (in case some elements r_{ij} are zero), convergence is ensured for ω between 0 and 2, $\gamma_j = \sum_i |r_{ij}|^\alpha$, $\rho_i = \sum_k |r_{ik}|^{2-\alpha}$, and $0 < \alpha < 2$.²⁶ (Other choices can also give convergence.)

Note that before pixel element j can be updated, all the paths i must have been processed so that the s_i^k values are available for all the paths and the summation over i can be evaluated. The same residual vector is used for one entire sweep through the N values of j . This is the main difference between ART and SIRT methods. ART updates the f_i after the equation for each path is treated. SIRT updates the f_i based on residuals averaged over all of the rows of R .

The special case $\alpha=0$, $\omega=1$ yields $\rho_i = \sum_j |r_{ij}|^2 = \langle R_i^T R_i \rangle$, $\gamma_j = \sum_i |r_{ij}|^0 = M_j =$ the number of paths that intercept pixel j , and

$$f_j^{k+1} = f_j^k + \frac{1}{M_j} \sum_i \frac{r_{ij} s_i^k}{\langle R_i^T R_i \rangle}.$$

This is the method used by Dines and Lytle when they probed the electromagnetic properties of the ground between boreholes.²⁷

Although Austen, Franke, and Liu presented the ART method,¹⁹ they used the SIRT method for their reconstructions. There they used $\alpha=0$ and $\omega=0.1$, and found diverging solutions for $\omega>0.3$. The magnitude of the noise affects the convergence properties of SIRT methods.

The special case $\alpha=1$, $\omega=1$ yields $\rho_i = \sum_k |r_{ik}| = L_i =$ the length of the propagation path i as it passes through the pixels, $r_j = \sum_i |r_{ij}| =$ the total length of all propagation paths passing through pixel j , and

$$f_j^{k+1} = f_j^k + \frac{1}{\sum_i |r_{ij}|} \sum_i \frac{r_{ij} s_i^k}{L_i}.$$

This is the method used by Hager et al.,²⁸ as noted in Reference 26.

Comparison with MART

The multiplicative algebraic reconstruction technique (MART) was adopted by Raymund et al.²⁰ in 1990 work at the University of Illinois. Rather than an iteration of the form $f_j^{k+1} = f_j^k + \alpha(T_j - T_j^k)$ as in ART, MART uses multiplicative modifications:

$$f_j^{k+1} = f_j^k \cdot \left(T_j / T_j^k \right)^\beta = f_j^k \cdot \left[\frac{T_j}{\langle R_i^T f^k \rangle} \right]^\beta,$$

where $0 < \beta < 1$. This approach reduces the tendency to converge toward any particular solution among all the possible solutions consistent with the known data element T_j . This property is desirable for cases where the data are incomplete; reconstructed images are less likely to introduce "ghosts" by the reconstruction process.

Earlier we used an initial guess $f_j^0=0$. With this choice, MART iterations would make little progress. An alternative approach would be to choose the $n_0(x,y)=0$ and let the f_j^0 be determined from the model of the ionosphere. Then, (with a

sufficiently high lower bound to the reconstruction region, h_1) all the f_j are real numbers greater than the interplanetary density of 100 electrons/cm³. The reconstruction area could still be restricted to the same area as before, perhaps with the T_i reduced by the contribution from propagation above the region being constructed. Since none of the f_j are zero or negative, incidence of "ghost" images should be less likely. However, treating curved paths would be more difficult.

After an initial guess of f^0 , the iteration procedure used the exponent $\beta = \omega_k r_{ij}$.¹⁹ At each path index i , the image f_j^k is updated to f_j^{k+1} . The inner product in the denominator might remain fixed until all the f_j^k have been updated for all j values, or the product might be altered after each increment in j . The $f_{j'}$ for j' values other than the subscript on the term f_j^k being processed (i.e. $j \neq j'$), only influence each iteration through this inner-product term. No smoothness is assumed. After the N pixels (j values) are processed for path index i , the i is then incremented to the next path. After the sweep through the $m \cdot n$ rows, convergence checks are performed and usually the next iteration starts through the i loop. The paths are organized in a random way to keep the algorithm from being biased to any particular paths or order of paths.²⁰

The exponent $\beta = \omega_k r_{ij}$ includes a relaxation parameter and a weighting by the length of the path-pixel intersection. A relatively small r_{ij} means that pixel has a small influence on the iteration. To keep the exponent dimensionless and less than 1, ω_k (in the interval 0,1) is divided by the maximum value of r_{ij} . The relaxation chosen by Raymund et al.²⁰ was 0.01.

MART and SIRT also have the advantage of versatility. The methods can be applied to incomplete problems, such as our ionospheric reconstruction. The chief advantage of MART (when applied to incomplete data problems) is its ability to avoid introducing "ghost images."

Alternate approach

UFO is a collection of interactive utility programs for estimating unknown functions of one variable using a wide class of information as input.²⁹ From its origin about three decades

ago, it was designed to help with ill-conditioned problems and with problems where the solution may not be unique. The programs allow constraints to be placed on the solution. For example, the f_j could be constrained to be positive and greater than 100 electrons/cm³. A set of neighboring pixels could be constrained to have a smooth variation in f_j . Tolerances can be inserted to limit the amount of "smoothness." Discontinuities can be built in to account for pixel-location jumps from row to row (or pixels might be renumbered so that pixels j and $j+1$ are always neighbors).

At first UFO was discarded as a possible analysis tool because the ionospheric reconstruction is in a two-dimensional (2D) region. However, the solution methods presented earlier all solve the matrix equation $Rf=T$, where the unknown vector, f , is a 1D column vector. The 2D features are included in the N pixel elements in f . In a sense, the earlier methods are treating the problem in 1D. Because of the convenient user interface and use on unfold problems in the past, we believe that UFO is a viable alternate method. An added convenience is the freedom to choose a variety of basis functions. All the earlier work on CT reconstruction of the ionosphere used pixel basis functions, that is, the electron densities were constant within a given pixel. The 1D UFO could easily allow variation with either the altitude or latitude inside each pixel. (It is not a true 2D treatment since the basis functions are limited to one dimension--altitude or latitude.) Iterative fittings could alternate between variation with altitude and variation with latitude within each pixel. If successful, the approach could be extended directly to reconstruction in a 3D region (at least for the simple "pixel" basis functions).

Example: Reconstruction from 800-km orbit and one transmitter

In our initial example, we choose a simple ionosphere model, calculate the resulting TEC data, and then attempt the inverse problem--that is, given the TEC data, generate the electron density profile.

The ionosphere is chosen as a step function approximation to a Chapman layer. The Chapman layer is given by

$$n_e = n_{\max} \exp[(1 - z - e^{-z})/2]$$

where $n_{\max} = 10^{12}$ electrons/ m^3 and $z = (h - h_0)/H_{\text{SCALE}}$. The Chapman layer is centered at altitude $h_0 = 350$ km, with a scale height of $H_{\text{SCALE}} = 75$ km. (A more general expression, useful for a larger area, would allow the parameters $n_{\max}$, h_0 , and H_{SCALE} to vary with the latitude and longitude.) The step function approximation has steps 100 km wide with seven steps spanning the region $100 \leq h \leq 800$ km. These distributions are illustrated in Figure 3.

Figure 3. Simple electron-density model for 100- to 800-km altitudes. Part b illustrates the independence from latitude in the example. The n_e units are electrons/ m^3 .

The satellite orbit and transmitter location must be identified to generate the TEC data. Here we select an 800-km orbit with measurements taken over a few hundred kilometers past the transmitter. The 18 paths that start from the single ground-based transmitter are illustrated in Figure 4. The ionosphere is divided into seven horizontal layers,

corresponding to the steps in Figure 3. Once the geometry is fixed, TEC data for the 18 paths and the tomography matrix can both be calculated for the 18 paths (rows) and the 7 pixels (columns). This process is automated by a computer code. The TEC values range from 2.98×10^{17} to 3.11×10^{17} as the path index varies from 1 to 18. The code also generates a problem definition file for directly inserting the tomography matrix and the TEC values into UFO.

Figure 4. Path and pixel geometry for single-transmitter example.

The UFO code finds the electron densities quickly, 4.26×10^{11} electrons/ m^3 in each of the 7 pixels. The fit gives an excellent match to TEC, but is not a good match to the profile in Figure 3! UFO also reveals that the tomography matrix has a rank deficiency of 6 (with only seven unknowns). Only one of the 18 paths is independent of the other 17. This results because the r_{ij} for a given path are the same for all pixels (j) and are a constant multiple of the r_{ij} for a

different path. Adding three other transmitters as in Figure 5 gives 72 paths, but offers no help. The rank deficiency remains at 6 and the pixel density remains at 4.26×10^{11} electrons/m³, as expected. The constant multiple is just the ratio of the slant TEC to the vertical TEC, which is the inverse of the cosine of the angle between the path and the vertical directions (for the horizontal layering).

Figure 5. Illustration of paths for a four-transmitter example.

The pixel divisions have been treated as "small" rectangles. This approximation resulted in the r_{ij} for a given path being the same for all pixels when the ionosphere was divided into the seven vertical layers. As an alternative, the pixel boundaries could be treated as vertical lines at the sides and arcs of circles at constant altitude for the top and bottom boundaries as indicated in Figure 1. This change makes the simulation more consistent with a Chapman layer and might lessen the rank deficiency.

A check of this hypothesis applied to the geometry of Figure 4 shows that subtracting the r_{ij} for path 1 (where all the values are 100 km) from all the r_{ij} then makes the $r_{ij}-100$ km for each row a constant multiple of the $r_{ij}-100$ km for row 2. The change to arc pixel boundaries does decrease the rank deficiency by one, but the matrix remains singular, giving two independent equations to determine the seven unknowns! The data remain incomplete. Specifically, the lack of data for rays parallel and nearly parallel to the surface of the earth inhibits determination of the vertical profile!

We suspect that, when path curvature is taken into account, the tomography matrix becomes nonsingular and the data become complete. However, a small amount of curvature will limit numerical realization of this advantage. The amount of curvature remains unknown until the electron-density profile and signal frequency are determined. Perhaps the curvature can help in an iterative scheme. An initial guess at the profile allows the curved paths to be calculated. The tomography matrix could then be determined. If not singular, the matrix and the TEC data could determine the electron-density profile and allow additional iterations until convergence.

We conclude that information must be added to the tomography data to determine the vertical profile. Let us try to determine what kind of information would be sufficient. Austen²¹ suggests using data from a satellite in another orbit or using sounder data to add nearly horizontal paths. Austen also finds that **when a reasonable initial guess is given, variations in the profile are reconstructed with his methods, even though the profile itself cannot be reconstructed.**²¹ Rather than implementing his techniques, let us examine other possibilities.

Suppose a Chapman shape were assumed. It then seems reasonable that the altitude of the peak intensity could be determined. The two independent equations should lead to a consistent estimate. With the h_0 known, then better estimates could be obtained for the other Chapman parameters, H_{SCALE} and n_{max} . The result would be a Chapman-layer fit to the data rather than a determination of the layering.

Suppose the peak electron density, its altitude, and the variation in density with altitude were known from ground level up to the peak (as perhaps from radar reflection data with a swept frequency). This, coupled with the TEC data, might be expected to help determine the vertical layering at higher altitudes. Unfortunately the earlier dilemma of a lack of independent equations remains for the vertical layering. Such attempts would be analogous to attempting the earlier reconstructions with a mesh that starts at the peak electron density and proceeds to higher altitudes. The TEC data are still not sufficient to determine the profile.

A similar reconstruction test that neglects the horizontal layers and assumes an artificial variation with latitude does detect the artificial layers. In this case, only one transmitter is sufficient to get independent paths, which resolve the horizontal structure and give no rank deficiency in the tomography matrix. Four transmitters reduce the uncertainties and improve the unfolded densities. These tests suggest that (1) vertical layers where the electron density varies only with latitude are easily reconstructed using our geometry of a receiver in a satellite with the transmitter(s) on the ground and (2) horizontal layers with electron density varying only with altitude require information added to the transmitter-to-satellite TEC data for successful reconstruction. The choices $n_H=4$ and $n_L=6$ as in Figure 1 still give a rank deficiency of 2.

Summary and Conclusions

A literature search verified that ionospheric structure is quite important. Its models are vital to understand several types of communication links to satellites. If identified, structures in the ionosphere can be included in signal analysis, and baud rates can increase.

The ionosphere reacts strongly to a series of geophysical processes. The examples described in this report are those most readily identified. Manmade environmental factors are changing. Nature is providing its own set of natural changes in the environment that leave their signature on the ionosphere. Monitoring our terrestrial environment requires a commitment to

record and archive data so the changes can be detected and correlations can be identified. Tomography offers a method to determine three-dimensional baseline data for ionospheric electron density, which is a portion of the information needed for prudent monitoring of our environment.

Several tomography approaches are described in this report. Methods applied to date have used high-frequency signals where straight-line paths are assumed. Even simple electron-density variations can be difficult to evaluate. Some limitations are identified and methods are illustrated. Despite difficulties, with reasonable constraints on the allowable ionosphere, not only is CT reconstruction of the ionosphere feasible, it has been accomplished with experimental data in at least two cases, by groups from Russia and from Wales. These cases were aided by considerable ionospheric variation with latitude. Reconstruction has also been demonstrated for simulated data in reports by groups from the University of Illinois, the Air Force Geophysics Laboratory, and by Northwest Research Associates in Bellevue, Washington.

Numerical tests indicate that the UFO code can be used for solution of computer-tomography problems. UFO has advantages of choices from a variety of basis functions and easy inclusion of tolerances and other constraints on the fit. In geometries that avoid singularities in the tomography matrix (either by insertion of constraints or by judicious choice of geometry), UFO is convenient to use.

Singularities in the tomography matrix are expected often. Our simple example of a horizontally layered ionosphere illustrated the problem that straight-line paths in rectangular pixels give only one independent equation regardless of how many transmitters and receiving stations are used. The alteration to pixels with curved upper and lower boundaries add a second independent equation. This difficulty makes the matrix singular and makes the solution less than unique. We expect that if the electron-density variation with altitude were represented by a summation over several orthogonal basis functions rather than by pixels having constant electron density, this limitation would remain. Only two coefficients that describe the variation with altitude alone can be uniquely determined. On the other hand, the variations with latitude are easier to establish, and are

better determined with added measurements. Constraints and/or variations with latitude are required to avoid singularities in the tomography matrix.

In the next part of this study of computerized tomography to determine ionospheric structures, we examine whether curved paths add further independent equations that help unfold the structure for simple, horizontally layered ionospheres.

References

- ¹M. C. Kelley, The Earth's Ionosphere, Academic Press, 1989.
- ²K. Davies, Ionospheric Radio, Peter Peregrinus Ltd, London, UK, 1990. This text replaces publications from 1965 and 1969. The 1965 work was to replace a 1948 publication of the (then) National Bureau of Standards, Circular 462.
- ³G. E. Backus, E. R. Benton, G. G. A. Harrison, and J. R. Heirtzler, The scientific case for magnetic field satellites, NASA-CR-186532, September 1987.
- ⁴Sunada Basu, S. Basu, E. MacKenzie, W. R. Coley, J. R. Sharber, and W. R. Hoegy, "Plasma structuring by the gradient-drift instability at high latitudes and comparison with velocity-shear driven processes," Jour.Geophys.Res., September 1970.
- ⁵J. M. Goodman and J. A. Klobuchar, "Ionospheric effects on communication and related systems (IES)," Radio Science 23:209(1988).
- ⁶Radio Science 20:259-574(1985).
- ⁷"Ionospheric Modification and its Potential to Enhance or Degrade the Performance of Military Systems," NATO Advisory Group for Aerospace Research and Development (AGARD-CP-485). Papers presented at Bergen, Norway, May 28-31, 1990. Available from NTIS, 5285 Port Royal Road, Springfield, Virginia 22161.

- 8S. Ganguly, "Effects of high power RF fields in the atmosphere and the ionosphere, Intense Microwave and Particle Beams," SPIE 1226:335-351(1990).
- 9S. Basu, "Ionospheric constraints of VHF/UHF communication links during solar maximum and minimum periods," Radio Science 23:363-78(1988).
- 10O. K. Garriott, A. V. da Rosa, W. J. Ross, "Electron content obtained from Faraday rotation and phase path length variations," J. Atmos. and Terrestrial Physics 32:705-27(1970).
- 11G. J. Bishop, J. A. Klobuchar, and P. H. Doherty, "Multipath effects on the determination of absolute ionospheric time delay from GPS signals," Radio Science 20:388-96(1985).
- 12R. S. Massey, A lookup table approach for the G/AIT algorithm, LA-UR-88-1439 (1988).
- 13R. S. Massey, J. Blake, and D. Hoard, Analysis of WSRP data using a lookup table approach: the VAWS algorithm, LA-CP-90-205, April 1990.
- 14J. Bacon, "Amateur Radio and Meteorology Go Hand-In-Hand," Practical Wireless, October 1990, pp 33-36.
- 15P. J. Edwards, "High-energy transient ionospheric monitoring network," Proc. of Astronomical Society of Australia, 8:263(1990).
- 16H. Rishbeth, "Ionospheric Science and Geomagnetism," Quarterly Journal of the Royal Society, 32:409-21(1991).
- 17J. Lastovicka, D. Pancheva, and X. J. Zhang, "On some difference in the behavior of the upper and lower part of the lower ionosphere in relation to meteorological parameters," Annales Geophysicae 8:599-608(1990).
- 18G. A. Hajj and T. P. Yunch, "The Possibility of 3-D Ionospheric Tomography with GPS," and Y. T. Chiu, R. M. Robinson, G. T. Davidson, and R. A. Rinaldi, "Simulation of Space-Based Ionospheric Tomography," Papers presented at the

1991 International Union of Radio Sciences, London, Ontario, Canada, June 24-28, 1991.

19J. R. Austen, S. J. Franke, and C. H. Liu, "Ionospheric imaging using computerized tomography," Radio Science 23:299-307(1988).

20T. D. Raymund, J. R. Austen, S. J. Franke, C. H. Liu, J. A. Klobuchar, and J. Stalker, "Application of computerized tomography to the investigation of ionospheric structures," Radio Science 25:771-89(1990).

21J. R. Austen, T. D. Raymund, C. H. Liu, S. J. Franke, J. A. Klobuchar, and J. Stalker, "Computerized ionospheric tomography," presented at the Ionospheric Effects Symposium, 1-3 May 1990.

22S. E. Pryse and L. Kersley, "An experimental determination of ionospheric tomography," 1991 North American Radio Science Meeting Program and Abstracts, p 551. The meeting was held at the University of Western Ontario, London, Ontario, Canada, June 24-28, 1991. Also, see I. K. Walker, L. Kersley, and C. D. Russell, "Mapping Electron Content Near the Main Trough," Proceedings of the International Conference on Antennas and Propagation, York, UK, 15-18 April, 1991, pp 906-909.

23E. S. Andreeva, V. E. Kunitsyn, Y. A. Melnichenko, and E. D. Tereshchenko, "The phase-difference ionospheric tomography," 1991 North American Radio Science Meeting Program and Abstracts, p 552. The meeting was held at the University of Western Ontario, London, Ontario, Canada, June 24-28, 1991.

24I. Koltracht, P. Lancaster, and D. Smith, "The structure of some matrices arising in tomography," Linear Algebra and its Applications, Elsevier Science Publishing Company, Inc., New York, 1990.

25F. Natterer, The Mathematics of Computerized Tomography, Wiley, New York, 1986, pp 128-138.

26A. van der Sluis and H. A. van der Vorst, "SIRT- and CG-Type Methods for the Iterative Solution of Sparse Linear Least-

Squares Problems," Linear Algebra and its Applications, Elsevier Science Publishing Co., New York, 1990, pp 257-281.

27K. A. Dines and R. J. Lytle, "Computerized geophysical tomography," Proc IEEE 67:1065-73(1979).

28B. H. Hager, R. W. Clayton, M. A. Richards, R. P. Comer, and A. M. Dziewonsky, "Lower mantle heterogeneity, dynamic topography and the geoid," Nature 313:541-45(1985).

29L. D. Kissel, F. Biggs, and T. Marking, UFO (UnFold Operator): Command Descriptions, SAND 82-0396, Sandia National Laboratories, June 1991.

Distribution:

J. Klobuchar
Phillips Laboratory (AFSC/PL)
Geophysics Directorate/LIS
Hanscom AFB, MA 01731-5000

Los Alamos National Laboratory
Attn: Paul Argo
Robert Massey
Tim Murphy
Tom Armstrong

G. W. Ulrich
Defense Nuclear Agency
6801 Telegraph Road
Alexandria, VA 22310-3398

J. M. Goodman
SRI International
1611 North Kent Street
Arlington, VA 22209-7192

J. Austin
Tennessee Technological University
Department of Physics
Cookeville, TN 38505

J. Thomason
Naval Research Laboratory
Radar Division (Code 5324)
Washington, DC 20375-5000

T. D. Raymund
Wave Propagation Laboratory
Department of ECE, UIUC
1308 W. Main Street
Urbana, IL 61801-2307

J. A. T. Heaton,
S. E. Pryse,
L. Kersley
University College of Wales

Department of Physics
Penglais
Aberystwyth, Dyfed SY23 3BZ, UK

E. J. Fremouw,
J. A. Secan
Northwest Research Associates, Inc.
Bellevue, WA 98009-3027

P. A. Bernhardt,
J. D. Huba,
P. K. Chaturvedi
Plasma Physics Division
Naval Research Laboratory
Washington, DC 20375-5000

A. J. Coster,
E. M. Gaposchkin,
L. E. Thornton
Millstone Radar/MIT Lincoln Laboratory
244 Wood Street
Lexington, MA 02173

1010	P. L. Mattern, LDRD Office
5902	W. D. Brown
5913	G. D. Cable
7141	Technical Library (5)
7151	Technical Publications
7613-2	Document Processing for DOE/OSTI (10)
8523-2	Central Technical Files
9206	N. Blocker
9221	D. E. Carroll
9225	A. Boye
9225	G. Elliott
9232	M. Sharp
9312	F. Biggs
9352	J. H. Renken
9352	T. L. Calton
9352	D. J. Riley
9352	C. N. Vittitoe (15)

END

DATE
FILMED

101 4 193