To appear in Radio Science, 2000

Incoherent scatter experiments at Jicamarca using alternating codes

D. L. Hysell

Department of Physics and Astronomy, Clemson University, Clemson, South Carolina

Abstract. An incoherent scatter experiment using alternating coded pulses has been implemented at Jicamarca. The experiment is intended to facilitate aeronomy research in the topside and protonosphere. By virtue of utilizing the full duty cycle capabilities of the radar transmitter, the experiment has improved sensitivity over the conventional double pulse experiment and permits multiple parameter estimation at protonospheric altitudes. Error analysis, including the estimation of the complete error covariance matrix for the alternating coded pulses, is discussed. Example density, temperature, and composition data illustrate the capabilities of the new experiment.

1. Introduction

Since it began operating in 1961, the Jicamarca Radio Observatory has been an important source of plasma density, temperature, composition, and drift measurements in the equatorial zone. Jicamarca's capabilities and observing modes have evolved over the years and are different from those of other incoherent scatter radars. (See *Farley* [1991] for a review of this evolution.) Its distinction is due partly to its proximity to the dip equator (11.95°S, 76.87°W, 1°N magnetic, 1°dip.) Jicamarca has a modular, phased array antenna with a main lobe that can be directed perpendicular to the geomagnetic field. The narrowness of the incoherent scatter spectrum for perpendicular backscatter permits very accurate measurements of cross-field plasma drifts, as demonstrated recently by *Kudeki et al.* [1999]. Perpendicular backscatter from intense, nonthermal, field-aligned irregularities in the *E* and *F* regions can also be observed at Jicamarca for the study of equatorial plasma instabilities. When its main beam is steered a few degrees off perpendicular, Jicamarca's crossed-dipole antenna can be used to measure the Faraday rotation of incoherent scatter, a technique which affords an independent, absolute estimate of electron density and provides a means of calibrating conventional power profiles.

Jicamarca is also unique in its operating frequency, 49.92 MHz. The sky noise temperature at this frequency varies between about 5000 and 45,000 K, depending on the sidereal time, limiting the radar sensitivity. Furthermore, the autocorrelation time of the incoherent scatter signal when observed a few tenths of a degree or more off perpendicular is of the order of 1 ms, and the autocorrelation function (ACF) must usually be measured to lags of $\sim$2 ms for accurate parameter estimation. This implies that the range resolution of a conventional long-pulse experiment can be no better than $\sim$300 km at Jicamarca for the shortest lags. Finally, coherent scatter from the equatorial electrojet received through the antenna sidelobes is an almost continuous source of clutter. The clutter contaminates signals from long-pulse experiments in range gates up to $\sim$450 km altitude, rendering them impractical for work near and below the *F* region peak.

Plasma temperature and composition measurements at Jicamarca have mainly been carried out using a double-pulse technique in which pairs of short ($\sim$100 μs) pulses are transmitted on opposing circular polarizations and with varying pulse spacings, ranging from 200 μs to 2 ms. The different lags of the ACF are then built up one at at time from the pulse pairs. Self clutter is nearly eliminated by the use of orthogonal polarizations, and clutter from the ground and from coherent echoes is mitigated by the experiment's low duty cycle. Only some lags of the ACF in some range gates are contaminated, and these are automatically rejected from the analysis. Many years' worth of temperature, density, and composition data collected with this technique are available for retrieval from the Web site at http://landau.phys.clemson.edu.

However, the sensitivity of the double-pulse experiment suffers from its very slow accumulation of data and from

its underutilization of Jicamarca's 6% maximum duty cycle. Measurement of parameters in the topside and protonosphere in particular require much greater sensitivity. A new coded long-pulse experiment was therefore implemented using the alternating codes introduced by *Lehtinen* [1986] [see also *Lehtinen and Häggström*, 1987; *Lehtinen and Huuskonen*, 1996] as an alternative to existing long-pulse, multipulse, and pseudorandom pulse techniques [*Farley*, 1972; *Holt et al.*, 1992; *Sulzer*, 1986]. Alternating codes are repeating sequences of binary phase codes that permit high resolution ACF measurements from long-pulse experiments. They differ from pulse compression codes such as Barker and complementary codes in that all of the lags of the ACF are derived from the time span of individual pulses. (Conversely, spectral information is derived from pulse to pulse measurements in Barker and complementary code experiments, which are therefore unsuitable for observing overspread targets.) Because long pulses are used, alternating codes permit the full exploitation of a radar's duty cycle. Furthermore, they provide multiple measurements of the lags of the ACF from each transmitted pulse and therefore foster very good statistics. The Lehtinen alternating codes are themselves derived from the Walsh sign sequences, the basis of an alternative form of fast Fourier transform. Their orthogonal nature permits the measurement of different lags of the ACF with range resolution determined by the baud length rather than the entire pulse length.

The ambiguity functions of Lehtinen alternating codes are localized in range and lag for each distinct lag product measurement. Cancellation of unwanted contributions to the ambiguity function is achieved systematically with alternating codes rather than statistically, which is the case for pseudorandom codes [*Sulzer*, 1986]. Two types of codes, weak and strong, have been found. Both have baud lengths that are a power of 2. Strong codes are necessary when of the bits of the code are transmitted continuously and without subcoding. For strong codes the number of pulses in a complete scan is twice the number of bits. Codes with lengths longer than 32 bits remained unknown for some time, but practical means of generating very long strong codes now exist [*Nygrén and Markkanen*, 1997; *Markkanen and Nygrén*, 1997]. Moreover, *Sulzer* [1993] developed a set of alternating codes that were not restricted to power of 2 lengths. However, ambiguity is achieved with these codes only with the summing of all the lag product measurements of a given lag. Such an approach becomes impractical if the transmitter power varies over the length of the pulse.

We will not dwell here on the theory of alternating codes, which are already in widespread use in the incoherent scatter community and which are explained in the references cited above. Rather, we discuss issues pertaining to their implementation at Jicamarca. We will mainly discuss error analysis and the estimation of statistical uncertainties. *Huuskonen and Lehtinen* [1996] address this issue and provide a recipe for estimating the error covariance matrix, a quantity required for parameter estimation by nonlinear least squares or generalized inverse methods. However, their method is an empirical one and requires the retention and storage of the lag product matrices associated with each coded pulse in the alternating code scan. Such a method is not expedient at Jicamarca, where data normally need to be transmitted over the internet for analysis by remote users. Furthermore, the theoretical formula is required for a maximum-likelihood formulation of the parameter estimation algorithm which would assign errors on the basis of the model ACF rather than the data. (Such an algorithm is planned but has not yet been implemented.) We describe below a theoretical approach to error analysis based solely on estimates of the ACF itself.

2. Data Analysis

The experiments in question employ a 16-bit strong Lehtinen alternating code with a pulse length of 2 ms and a baud length of 125 μs, representing a favorable compromise between sensitivity and range resolution. The interpulse period is 40 ms. The pulse pattern provides for 18.75 km range resolution and for 15 nonzero lag measurements at integer multiples of 125 μs. The data are decoded using matched filtering, and the autocorrelation functions are computed from them for each range gate. Summing of the ACFs takes place over the entire 32-pulse scan. Because the ambiguity function associated with the complete series of pulses is highly peaked and localized in range and lag, the ACFs for different range gates are measured distinctly for all but the zero lag. For parameter estimation we make no use of measurements of the zero lag but instead treat it as one more parameter to fit. The zero lag power profile does, however, contribute to a satisfactory estimate of the electron density profile and is also useful for error analysis (see below). Normalization of the power profile is carried out using ionosonde data. Line-of-sight plasma drifts at the dip equator can be measured much more accurately using techniques other than the long pulse, and we will not attempt to make such measurements here. Consequently, we will ignore the imaginary part of the ACF estimates throughout our analysis and assume that the actual ACFs are purely real.

Having computed estimates of the ACF for each range gate in this manner, we must assign statistical confidences before proceeding on to parameter estimation. Statistical inverse theory in general and nonlinear least squares techniques in particular can be carried out only with an estimate

of the error covariance matrix $\mathbf{C}$, which has the elements $C_{\tau\tau'} \equiv \langle \mathrm{Re}[\rho(\tau) - \hat{\rho}(\tau)]\mathrm{Re}[\rho(\tau') - \hat{\rho}(\tau')]\rangle$, in which $\rho(\tau)$ refers to the measured estimate of the ACF for lag τ, where the caret denotes the expected value, and where the angle brackets represent the ensemble average. (In practice, the ensemble average is replaced by a time average.) With the statistical errors that cause the estimates to depart from their expected values being jointly normal Gaussian random variables in accordance with the central limit theorem, we can express the probability of measuring a given ACF as

$$P(\underline{\rho}) = \frac{1}{(2\pi)^{\frac{n}{2}}|\mathbf{C}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}\left(\underline{\rho} - \underline{\hat{\rho}}\right)^t \mathbf{C}^{-1}\left(\underline{\rho} - \underline{\hat{\rho}}\right)\right] \tag{1}$$

where the column and row vectors have different components for different lags of the ACF. Now, the expected values of $\underline{\hat{\rho}}$ are complicated nonlinear functions of parameters including temperature, density, and composition, which can be represented by a vector $\underline{p}$. However, if we expand that functional dependence about an initial guess $\underline{p}_\circ$ as

$$\underline{\hat{\rho}}(\underline{p}) \quad = \quad \underline{\hat{\rho}}(\underline{p}_\circ) + (\underline{p} - \underline{p}_\circ)\frac{\partial \underline{\hat{\rho}}}{\partial \underline{p}} + \cdots \tag{2}$$

and truncate the series at the linear term, then it can be shown that the probability in (1) is maximized by the parameters given by

$$\underline{p} - \underline{p}_\circ \quad = \quad \left(\mathbf{J}^t\mathbf{C}^{-1}\mathbf{J}\right)^{-1}\mathbf{J}^t\mathbf{C}^{-1}[\underline{\rho} - \underline{\hat{\rho}}(\underline{p}_\circ)], \tag{3}$$

where $\mathbf{J}$ is the first derivative matrix in (2) [e.g., *Bevington*, 1969]. The nonlinear least squares technique then involves iterating (3) to convergence. The uncertainty for the ith final parameter estimate is then given by the transformation:

$$\sigma_i^2 \quad = \quad \left(\mathbf{J}^t\mathbf{C}^{-1}\mathbf{J}\right)^{-1}_{ii}. \tag{4}$$

2.1. General Error Covariance Analysis

Both the radar signal scattered from a given scattering volume detected at the receiver and the statistical errors associated with our ACF estimators are multivariate Gaussian random variables characterized by their covariances. Consider the general problem of the error covariance of two correlation measurements made repeatedly from four samples or a Gaussian random process at times t, $t + \tau$, t', and $t' + \tau'$. Label the ith realization of these four samples as $\{V_{1i}, V_{2i}, V_{3i}, V_{4i}\}$. Further, define an estimate of the correlation function to be

$$\rho_{12} \quad = \quad \frac{1}{Sk}\sum_{i=1}^{k} V_{1i}V_{2i}^*, \tag{5}$$

where S is the variance of the complex signal or the signal power and k is the number of statistically independent samples. We will neglect biases associated with this estimator, assuming that k is large enough to render them negligible. As stated by *Huuskonen and Lehtinen* [1996] and as can easily be shown by applying the formalism of *Farley* [1969],

$$\begin{aligned} C_{\tau,\tau'} \quad &\equiv \quad \langle \mathrm{Re}(\rho_{12} - \hat{\rho}_{12})\mathrm{Re}(\rho_{34} - \hat{\rho}_{34})\rangle \\ &= \quad \frac{1}{2k}\mathrm{Re}\left(\hat{\rho}_{13}\hat{\rho}_{24}^* + \hat{\rho}_{23}\hat{\rho}_{14}^*\right) \end{aligned} \tag{6}$$

The result is that the error covariance does not depend explicitly on the expectation of $\rho(\tau)$ or $\rho(\tau')$ but rather on the expectation of the ACF evaluated at the other four possible time permutation. Since radar experiments involve samples made at regular intervals, however, $\hat{\rho}(\tau)$ and $\hat{\rho}(\tau')$ frequently reenter the picture. The error covariances $C_{\tau,\tau'}$ for four illustrative situations are given below:

$$\begin{aligned} t = t' \qquad & \frac{1}{2k}\mathrm{Re}\left[\hat{\rho}^*(\tau' - \tau) + \hat{\rho}^*(\tau)\hat{\rho}^*(\tau')\right], \\ t + \tau = t' \qquad & \frac{1}{2k}\mathrm{Re}\left[\hat{\rho}(\tau)\hat{\rho}^*(\tau') + \hat{\rho}^*(\tau + \tau')\right], \\ t + \tau = t' + \tau' \qquad & \frac{1}{2k}\mathrm{Re}\left[\hat{\rho}(\tau - \tau') + \hat{\rho}^*(\tau')\hat{\rho}^*(\tau)\right], \\ t = t', \tau = \tau' \qquad & \frac{1}{2k}\mathrm{Re}\left[1 + \hat{\rho}^{*2}(\tau)\right]. \end{aligned}$$

The use of common samples in the computation of $\hat{\rho}(\tau)$ and $\hat{\rho}(\tau')$ introduces a factor of the zero lag $\hat{\rho}(0) = 1$ into the first three formulas. The last formula is an error autocovariance and contains two factors of the zero lag, which appear as the unity term.

Note that alternating code experiments involve data from multiple scans, each composed of multiple phase codes. The individual phase codes may give rise to samples from different statistical populations with different lag product expectations. Only when all the lag products are added will the desired expectations and ambiguity be achieved. The appropriate generalization for (6), taking into account samples from different phase codes and different populations, is

$$C_{\tau,\tau'} \quad = \quad \frac{1}{2k}\mathrm{Re}\left(\langle\hat{\rho}_{13}\hat{\rho}_{24}^*\rangle + \langle\hat{\rho}_{23}\hat{\rho}_{14}^*\rangle\right), \tag{7}$$

where the angle brackets denote an average over all pulses comprising a scan and where k is now the number of scans times the number of pulses in a scan. In the event that the expectations of given lag products for different pulses in a scan have small or uncorrelated deviations, $\langle\hat{\rho}_{13}\hat{\rho}_{24}^*\rangle \rightarrow \langle\hat{\rho}_{13}\rangle\langle\hat{\rho}_{24}^*\rangle$ (for example), and the lags on the right side of the arrow have the desired expectations and ambiguity. Throughout section 2.2 we will drop the angle brackets and assume that the properties of the coded pulses are such that

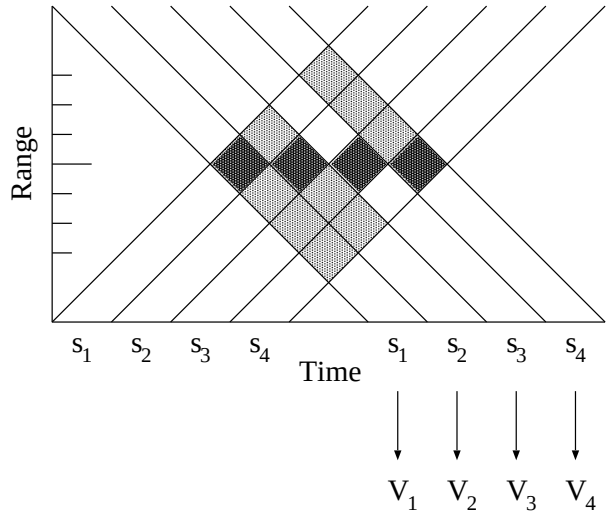

Figure 1. Range-time diagram illustrating a hypothetical 4-bit coded long-pulse experiment. In this example an estimate of the first (second) lag of the ACF can be computed from the samples denoted V_0 and V_1 (V_1 and V_3). A significant error covariance will arise from the shared use of the V_1 sample. Dark shaded regions represent signal, while light shaded regions represent clutter.

decorrelation holds. In fact, this is only approximately true for so-called randomized alternating codes and untrue for traditional alternating codes. The discrepancy will be addressed in section 2.3.

2.2. Effects of Clutter and Noise

The preceding arguments hold generically but need to be modified for practical application to the alternating code experiment. Below, we discuss the modifications that need to be made, following an approach similar to the one outlined by *Farley* [1969]. Figure 1 depicts a range-time diagram for a hypothetical alternating code experiment using a 4-bit code sequence. The s_1 ... s_4 refer to the signs of a binary code sequence employed upon transmission and during decoding. The shaded regions in the diagram represent backscatter arising from the range interval of interest. When the data are properly decoded, the nonzero lags of the ACF computed from the samples V_1 ... V_4 will reflect properties of the scattering medium in the indicated range interval only. The correct ACF estimator should therefore be

$$\rho_{mn} \quad = \quad \frac{1}{Sk}\sum_{i=1}^{k} V_{mi}V_{ni}^*, \tag{8}$$

where the sample voltages have been decoded and where S represents the signal power arising from the range gate of interest only. However, this S is no longer the variance of the signal, which now contains sky noise and also clutter from undesired range gates. We therefore rewrite (8) as

$$\rho_{mn} \quad = \quad \frac{[S+N+C]_{mn}}{S}\frac{1}{[S+N+C]_{mn}}\frac{1}{k}\sum_{i=1}^{k} V_{mi}V_{ni}^*, \tag{9}$$

where N and C represent the noise and clutter power, respectively. The square brackets denote the geometric mean of the signal plus noise plus clutter powers associated with the samples m and n; since the scattering process is nonstationary, the clutter power will, in fact, be different for different samples used to calculate the ACF for a given range gate. At Jicamarca, sky noise dominates noise in the receivers and has a strong diurnal variation. The clutter, meanwhile, will always be much stronger than the signal, by a factor something like the baud length of the code less one, and will be the limiting factor on the sensitivity of the experiment up to very high altitudes at Jicamarca.

Now the factors in (9) to the right of the first quotient represent an autocorrelation function measurement normalized to the variance of the signal, precisely what we have expressed generally in (5). This means that the expressions derived above in (6) and thereafter are applicable to a coded long-pulse experiment provided that they are multiplied by pairs of factors of $[S+N+C]_{mn}/S$ with the appropriate subscripts. However, the interpretation of the results is now somewhat different, as the noise and clutter portions of the signal are uncorrelated in all but the zero lag. Estimates of the ACF will therefore be reduced by the factor

$$\hat{\rho}_{mn} \quad = \quad \hat{\rho}_{smn}\frac{S}{[S+N+C]_{mn}}, \tag{10}$$

where $\hat{\rho}_s$ represents the ACF of that portion of the signal arising from backscatter from the desired range gate. It is this quantity in which we are interested. We assume here that the receiver filters are perfectly matched. If not, then even the strong code will give rise to correlated clutter that not only will influence the error covariances but could induce a bias in the ACF estimates. This concern is, however, beyond the scope of the present analysis. Let us therefore substitute (10) into (6), multiply by two corresponding factors of $[S+N+C]_{mn}/S$, solve in terms of ρ_s, and henceforth drop the s subscript. The result is an error covariance estimator similar to (6) except in one crucial respect. That is, wherever a zero lag is implied in (6), the formula in (10) should not be used. The signal, noise, and clutter power all contribute to the zero lag estimator, and $\hat{\rho} = 1$ given the normalization in use.

We revisit below the example error covariance estimates given previously, now written more explicitly for an alternating code experiment:

$$\begin{aligned}
t = t' \qquad & \frac{1}{2k}\mathrm{Re}[\hat{\rho}^*(\tau' - \tau)R_m + \hat{\rho}^*(\tau)\hat{\rho}^*(\tau')]\,, \\
t + \tau = t' \qquad & \frac{1}{2k}\mathrm{Re}[\hat{\rho}(\tau)\hat{\rho}^*(\tau') + \hat{\rho}^*(\tau + \tau')R_m]\,, \\
t + \tau = t' + \tau' \qquad & \frac{1}{2k}\mathrm{Re}[\hat{\rho}(\tau - \tau')R_m + \hat{\rho}^*(\tau')\hat{\rho}^*(\tau)]\,, \\
t = t', \tau = \tau' \qquad & \frac{1}{2k}\mathrm{Re}\left[R_m R_n + \hat{\rho}^{*2}(\tau)\right]\,.
\end{aligned}$$

where R_m is the ratio $(S + N + C)/S$ for the sample m. The geometric mean notation is gone since the R factor survives only where common samples are in use. Error covariances of ACF measurements that involve common samples will therefore be much larger than those that do not, particularly in the low signal-to-noise case but always as a rule. Autocovariances will be larger still, all the samples being used in common in that case.

2.3. Correlated Clutter

In this section we revisit the assumption made following (7) that the expectations of given lag products associated with different pulses comprising a scan have small or uncorrelated deviations. This assumption is, in general, unjustified. Whereas the codes have been designed such that clutter is exactly canceled and does not appear in the first moment averages $\langle\hat{\rho}\rangle$, clutter can and generally will contribute significantly to the second moment terms $\langle\hat{\rho}\hat{\rho}^*\rangle$. Consequently, the particular construction of a given set of codes affects the speed of the radar experiment; even codes with the same ambiguity function can have different speeds.

Note, however, that our earlier results are unaffected in the case that one or both of the factors in the second moment terms is a constant, namely, R. The only covariance terms that are affected by correlated clutter therefore are the ones that arise from cases where no samples are shared between lags. Even though these terms outnumber terms where samples are shared, they might otherwise be neglected in view of the fact that R is much greater than unity. These terms must be considered, however, whenever correlated clutter is an issue and can, in fact, contribute significantly to the experimental uncertainty.

Lehtinen et al. [1997] raise the issue of correlated clutter and investigate the relative speed of alternating codes, random codes, and randomized codes. They demonstrate that some coding schemes perform better than others and argue that the effects of correlated clutter are minimized by using randomized alternating code sequences. Consider that much of the correlated clutter in a traditional Lehtinen alternating code sequence arises from the first code which is made up entirely of bauds with the same sign. Randomizing codes alleviates this problem and generally reduces the size of the second moment terms to the point that they no longer compete with the terms of order R or R^2.

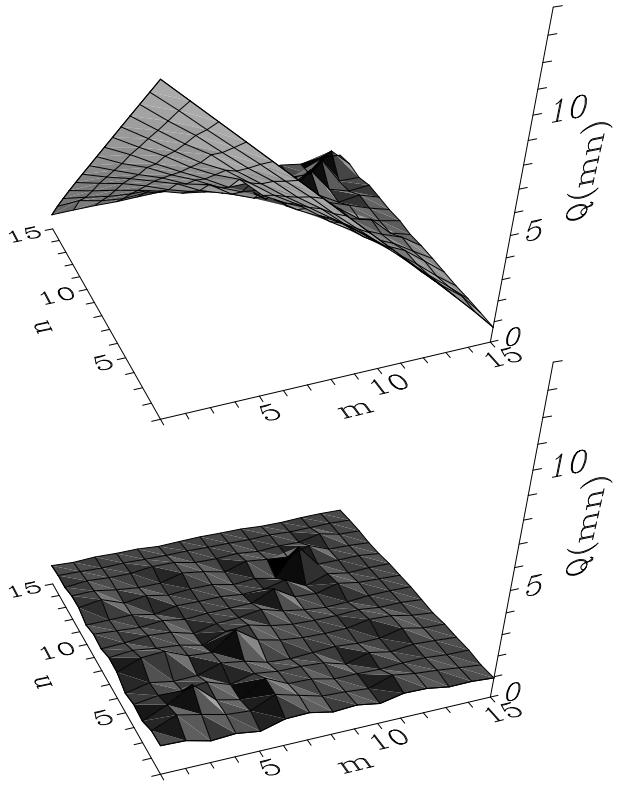

Figure 2. Simulated effect of correlated clutter associated with a randomized alternating code (bottom) and the traditional alternating code (top). Here the normalized second moment of the ACF is plotted against the lag number.

A Monte Carlo computer algorithm has been used to estimate the size of Re $\langle\hat{\rho}(\tau)\hat{\rho}^*(\tau')\rangle$ for the 16-bit strong code. The algorithm involves generating multivariate Gaussian random variables representing the scattered signal components corresponding to each box in Figure 1 (or rather a 16 baud version of Figure 1). The random variables are generated so as to be statistically independent from range to range and to have ACFs representative of the incoherent scatter signal at Jicamarca. Alternating code sequences were used to code and decode the simulated signal. The simulation was then run for all the possible permutations of τ and τ'.

Figure 2 presents the results of the Monte Carlo simulation. It shows two-dimensional plots of the quotient $Q_{t,t'} = \langle\hat{\rho}(\tau)\hat{\rho}(\tau')\rangle / \langle\hat{\rho}(\tau)\rangle\langle\hat{\rho}(\tau')\rangle$, representing the effect of correlated clutter on the second moment of the ACFs de-

rived from individual coded pulses. The abscissas of the relief plots are the integer lag numbers corresponding to τ and τ' and ranging from 1 to 15. There are multiple ways of measuring a given lag and a multiplicity of ways of forming a given second moment; what is plotted here is the average computed from all the ways of estimating the second moment in question. The various terms contributing to the averages, in fact, deviate from them only slightly.

Most of the correlated clutter in the traditional experiment (top half of Figure 2) derives from the first pulse in the scan, which has constant phase, and the 22nd pulse, which simply alternates in sign (i.e., +-+-....). Consequently, we find that the symmetric matrix Q_{mn} is well approximated by the simple expression $Q_{mn} \approx 2(N-m)(N-n)/2N$, where N is the baud length of the code. (A better approximation is given by $Q^2_{mn} \approx [(N-m)(N-n)/N]^2+1$ which decays to unity for long lags.) This formula can henceforth be incorporated in the error analysis of data from traditional 16-baud Lehtinen alternating code experiments.

An intuitive picture of the effect of correlated clutter on the experimental statistics in the high signal-to-noise limit becomes apparent when assessing the experiment in the frequency domain (M. Sulzer, personal communication, 2000). The objective of pulse coding may be seen as widening the spectra of signals from unwanted ranges in such a way that, when added across a complete scan, they form a flat noise spectrum (the clutter spectrum). In effect, coding attempts to turn the clutter into a random process. However, when a particularly unsuitable code such as one with constant phase is transmitted, no broadening occurs, and the clutter from that pulse remains concentrated at low frequencies, in the interesting portion of the spectrum. Clutter from other pulses in the scan must then fall mainly outside the interesting portion of the spectrum in order to achieve a flat clutter spectrum overall. This has the effect of reducing the number of independent clutter samples in the low-frequency portion of the spectrum from what could otherwise be achieved with more optimal phase codes and a more uniform spectral redistribution of clutter power. Fewer independent clutter samples degrade the statistics of the clutter spectrum and degrade the experiment overall in the high signal-to-noise limit.

For completeness' sake, we repeated the Q_{mn} simulation for a randomized code experiment. This time, all the codes in a scan were multiplied by one of four random patterns that were then rotated from scan to scan. The particular random patterns were the ones tabulated and investigated by *Lehtinen et al.* [1997]. The results are shown in the bottom half of Figure 2. Evidently, evidence of correlated clutter has all but vanished in this case, as Q_{mn} has been reduced essentially to unity. The benefits of code randomization are therefore clear, and randomization schemes more elaborate than the one employed here are unnecessary. The radar controller at Jicamarca is currently undergoing an upgrade that will permit the transmission of randomized alternating codes. The upgrade will reduce covariances in the high signal-to-noise limit and will also supress range correlations associated with the correlated clutter. The data presented here, however, were all taken using the traditional alternating codes and consequently suffer from somewhat degraded statistics, particularly in the short lags in the high signal-to-noise regime. *Lehtinen et al.* [1997] have shown that the autocovariances in this regime can be reduced by about one third by code randomization.

Finally, we revisit one last time our example error covariance estimates, now including the effects of correlated clutter:

$$
\begin{aligned}
&t = t' \\
&\qquad \frac{1}{2k}\mathrm{Re}\,\left[\hat{\rho}^*(\tau'-\tau)R_m + \hat{\rho}^*(\tau)\hat{\rho}^*(\tau')Q_{\tau,\tau'}\right], \\
&t+\tau = t' \\
&\qquad \frac{1}{2k}\mathrm{Re}\,\left[\hat{\rho}(\tau)\hat{\rho}^*(\tau')Q_{\tau,\tau'} + \hat{\rho}^*(\tau+\tau')R_m\right], \\
&t+\tau = t'+\tau' \\
&\qquad \frac{1}{2k}\mathrm{Re}\,\left[\hat{\rho}(\tau-\tau')R_m + \hat{\rho}^*(\tau')\hat{\rho}^*(\tau)Q_{\tau,\tau'}\right], \\
&t = t', \tau = \tau' \\
&\qquad \frac{1}{2k}\mathrm{Re}\,\left[R_m R_n + \hat{\rho}^{*2}(\tau)Q_{\tau,\tau}\right].
\end{aligned}
$$

Here the carets once again refer to the expectations of a given lag of the ACF after summing across entire scans (i.e. the angle brackets have again been dropped).

2.4. Summary Analysis

What we have investigated so far is the error covariance between estimates of individual lags of the ACF. Coded long-pulse experiments provide $l = n - m$ means of estimating the mth lag of the ACF, where n is the baud length of the code. The complete error covariance matrix must express the covariances of all of the possible estimates of a given lag of the ACF with all of the possible estimates of another. The final expression then becomes

$$C_{\tau,\tau'} \quad = \quad \frac{1}{l}\frac{1}{l'}\sum_{i=1,j=1}^{l,l'} C_{\tau_i,\tau'_j}, \tag{11}$$

where l and l' are the number of ways of making estimates of the lags τ and τ', respectively, from the coded pulse. Each of the summed terms on the right side of (11) is formed in the manner of the examples shown above. It is particularly important to note that the error autocovariances too must be

calculated according to this formula and that the cross terms cannot be neglected. That is,

$$C_{\tau,\tau} = \frac{1}{l}\langle C_{\tau_i,\tau_i}\rangle + \frac{2}{l^2}\sum_{i=1,j>i}^{l} C_{\tau_i,\tau_j} \qquad (12)$$
$$> \frac{1}{l}\langle C_{\tau_i,\tau_i}\rangle,$$

where the second term in the sum represents the correlation between errors in the ith measurement of the ACF at a given lag with errors in the jth measurement. Evidently, the error autocovariances do not decrease simply as $1/l$ in coded long-pulse experiments as they do in many other contexts, as neglecting the cross correlations gives rise to a significant underestimate.

The error autocovariances will generally be much larger than the off-diagonal terms, however, making the covariance matrix diagonally dominant. It should be permissible in many instances to neglect the off-diagonal terms throughout the data analysis and parameter estimation, but they should not be neglected when determining the error bars on the fit parameters according to (4). *Huuskonen and Lehtinen* [1996] perform an analysis showing that neglecting these terms leads to an underestimate of the size of the error bars by up to a factor of ∼1.5.

2.5. Signal, Noise, and Clutter Estimates

The portion of the radar signal corresponding to sky and system noise is a stationary random variable, and we can estimate the noise power in every range gate from samples corresponding to high altitudes from which little backscatter returns. We presume that the expectation of the zero lag alone is affected by noise and estimate the noise level from it accordingly. The backscatter signal, meanwhile, is a nonstationary random process. The signal and clutter power estimates required by the preceding analysis must be representative of the scattering volume defined by the coded pulse ambiguity function, making the zero lag data an unsuitable basis overall. (We can, however, derive very robust if uncoded estimates of the electron density profile from the zero lag data with the caveat that they will be somewhat inaccurate below the *F* peak where the density changes rapidly with altitude. In that region the real part of the first lag of the ACF provides a better basis for the estimate of n_e.)

A number ways of estimating signal power profiles present themselves. Short uncoded pulses could be interspersed with the alternating coded pulses, and the power profiles could be derived from them. Such a method will be implemented in the near future when the alternating code and Faraday double pulse experiments are integrated at Jicamarca. Alternatively, the signal power could be treated self-consistently during parameter estimation. The zero lag, or, equivalently, the normalization, of the ACF already appears as an additional fit parameter in our analysis. The signal power and electron density profiles could therefore be expressed implicitly in terms of this parameter during the fitting. In the present experiments we use the first lag of the ACF as a proxy for the signal power at altitudes near and below the *F* peak, where the electron density changes rapidly but the composition is constant. Above that we use the uncoded zero lag divided by the number of bits in the alternating code. This approximation holds where the electron density changes slowly with altitude.

3. Experimental Results

In practice, we perform parameter estimation utilizing a Levenberg-Marquardt algorithm rather than the nonlinear least squares iteration technique outlined in section 2. Levenberg-Marquardt is used to minimize the chi-square deviation between model- and data-derived ACFs by employing a globally stable hybrid conjugate gradient/Newton iteration method. The most expedient application of the method involves diagonalizing the inverse covariance matrix found above. Then the argument of (1) can be written simply as

$$-\frac{\chi^2}{2} = -\sum_i \frac{(\tilde{\underline{\rho}} - \hat{\tilde{\underline{\rho}}}(\underline{p}))_i^2}{2\sigma_i^2}, \qquad (13)$$

where the tildes indicate vectors that have undergone a similarity transformation using the eigenvectors of $\mathbf{C}^{-1}$, the σ_i^2 are the inverses of the eigenvalues of $\mathbf{C}^{-1}$, and the sum is over the components of the vector $\tilde{\underline{\rho}} - \hat{\tilde{\underline{\rho}}}$.

We present here two examples of data taken at Jicamarca and processed according to the prescriptions given above. In Figure 3 are shown sample data taken after sunset on October 8, 1999. The data were taken using a single transmitter which generated a peak power of ∼1 MW and which drove the antenna array in circular polarization. The antenna itself was configured in what has historically been called its 4.5°position. In this configuration the main beam of the antenna is directed precisely 3.15°, 2.79°, and 2.44°off perpendicular to $\mathbf{B}$ at altitudes of 600, 900, and 1200 km, respectively.

The first column shows measured ACFs versus altitude in the form of error bars. The error bars themselves reflect autocovariances calculated according to (12). The solid lines drawn through the error bars are the best fit models. Electrojet clutter contaminated data below ∼450 km altitude (not shown). The incoherent integration time for these data was ∼25 min, although data were rejected a considerable amount of this time because of satellite contamination.

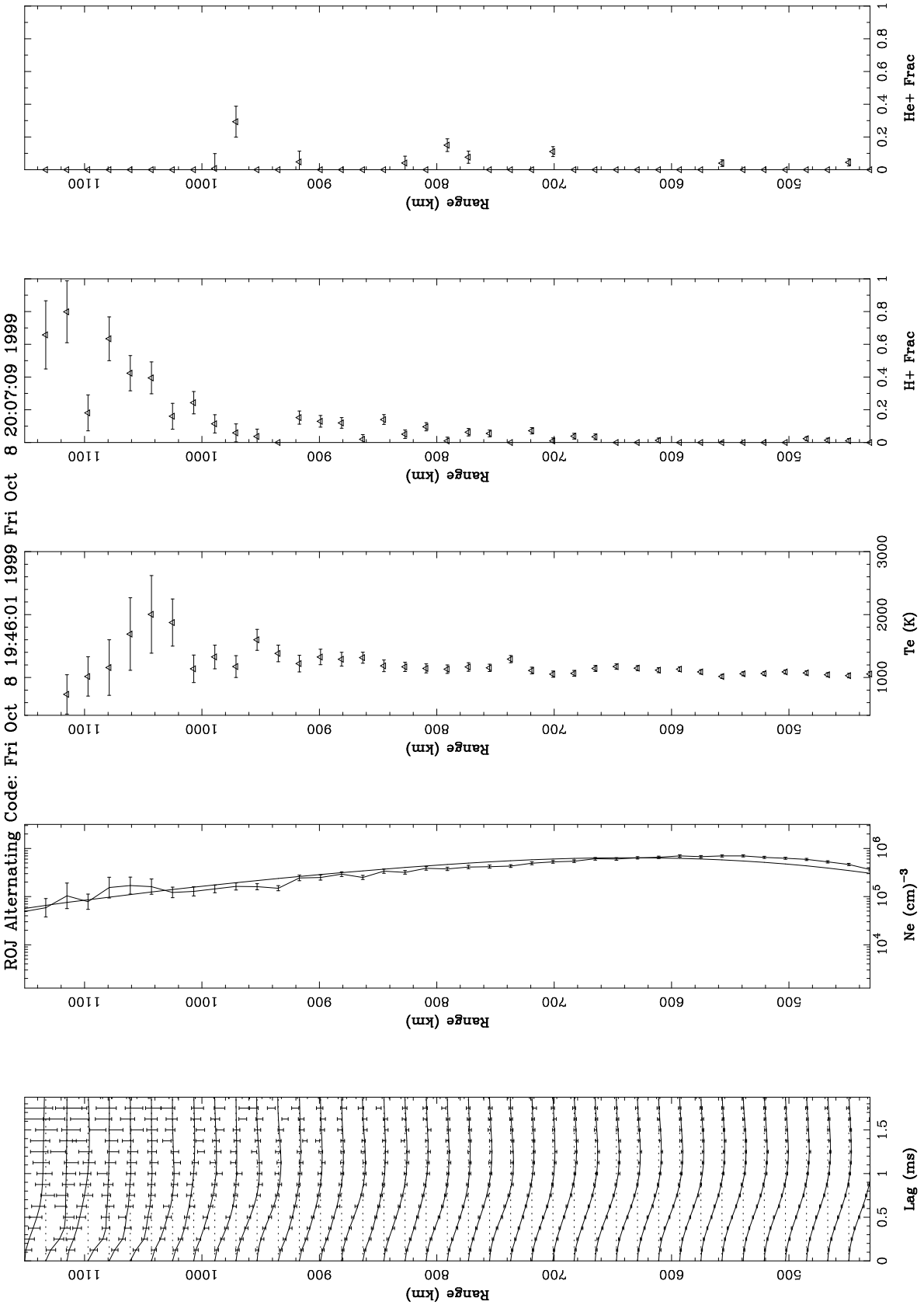

Figure 3. Alternating code incoherent scatter data for October 8, 1999.

The second column shows the magnitude of the measured zero lag (solid line) and of the best fit zero lag arising from the parameter estimation (line with error bars). Whereas the range resolution of the first of these curves is limited by the pulse length, that of the second is limited only by the baud length. The directly measured curve is, however, more robust, the signal-to-noise ratio for the zero lag being greater than the others by approximately a factor of the baud length of the code. Both curves have been range corrected, normalized to ionosonde measurements of $f_o f_2$, and plotted against an absolute scale of electron density. Background noise was estimated from the highest few altitudes and subtracted from the measured zero lag estimate prior to scaling and normalization. Note that $T_e \approx T_i$ at the altitudes and local time in question and that no temperature ratio correction has been performed.

The remaining three panels depict the electron temperature, hydrogen fraction, and helium fraction, as derived from the ACFs. We have assumed $T_e = T_i$ here and so fit only for these quantities and for the zero lag itself. The error bars were derived following (4). Clearly, data from the alternating code experiment at Jicamarca are of sufficiently high quality to permit multiple parameter estimation up to protonospheric altitudes. By comparison, given comparable power levels and integration times, the usual Faraday double pulse experiment at Jicamarca does not perform well above about 600-800 km, does not provide usable hydrogen fits above the *F* peak, and does not yield useful helium estimates at all.

It should be mentioned that temperature measurements at Jicamarca are historically problematic for reasons that are only now becoming clear. Measurements of T_e/T_i have long yielded values significantly less than unity with no apparent geophysical justification. The effect diminishes as the main beam of the Jicamarca antenna is steered further from perpendicular to $\mathbf{B}$ [*Pingree*, 1990]. Altitudes immediately above the *F* peak were the most affected. For many years it was suspected that light ion concentrations were responsible for the anomalously low electron temperature. However, this hypothesis could not readily be verified on the basis of conventional double pulse data because of the poor sensitivity of that experiment. Recently, *Sulzer and Gonzalez* [1999] argued that the effect is instead due to electron coulomb collisions, something normally not accounted for in the computation of electron admittance functions. Evidence supporting their hypothesis is mounting, but a practical means of accounting for the effects of electron coulomb collisions in the data analysis remains to be formulated. We have attempted to sidestep the issue to some extent here by focusing on post-sunset observations and enforcing the T_e=T_i constraint in our analysis. However, our admittance functions nevertheless require correction and will be updated once an expedient algorithm is available.

A data set similar to the one in Figure 3 but for November 11, 1999, is shown in Figure 4. This time, two of Jicamarca's megawatt transmitters were combined through a hybrid network and used to drive one circular polarization of the antenna array together. The extra sensitivity this afforded permitted good results to be obtained to ∼200 km higher altitude than before. Here we see that useful ACF and parameter estimates are available throughout the topside and well into the protonosphere.

4. Summary

An alternating code incoherent scatter experiment has been implemented at Jicamarca for the purpose of aeronomy research in the topside and protonosphere. The experiment improves upon the performance of the standard double-pulse mode at altitudes above ∼450 km by fully utilizing the data cycle capabilities of the Jicamarca transmitters. An approximate method for performing error analysis has been outlined that requires nothing more than estimates of the ACF, the electron density profile, and signal-to-noise and clutter-to-noise ratios for every range gate. Lag product matrices for individual pulses need not be retained to perform the analysis.

In the future, we plan to combine the alternating code and double-pulse experiments into a single experiment. The Faraday double-pulse technique provides absolute estimates of the electron density profile and yields measurements of the ACF down to altitudes just above the valley region. (Note that we presently cannot use the alternating code experiment effectively whenever the *F* peak falls below 450 km and power profile normalization becomes impossible). Parameter estimation from double-pulse data is sufficiently accurate in altitudes near the *F* peak, where the signal-to-noise ratio is high. The double-pulse experiment may even statistically outperform the alternating code experiment there since the former is not clutter limited. Error analysis for the double-pulse experiment is straightforward since the statistical errors associated with each lag of the ACF are uncorrelated.

Finally, we plan to attempt to extract line-of-sight drifts from the information in the complex ACF measurements. Obvious generalizations of the formulas presented here exist for the complex case as well. Plasma density, temperature, composition, and drifts could thereby be measured simultaneously at Jicamarca for the first time.

Acknowledgments. The author is thankful to M. Sulzer and S. Gonzales for their advice and for sharing their Jicamarca data and to K. Yegin for his assistance. This work was supported by

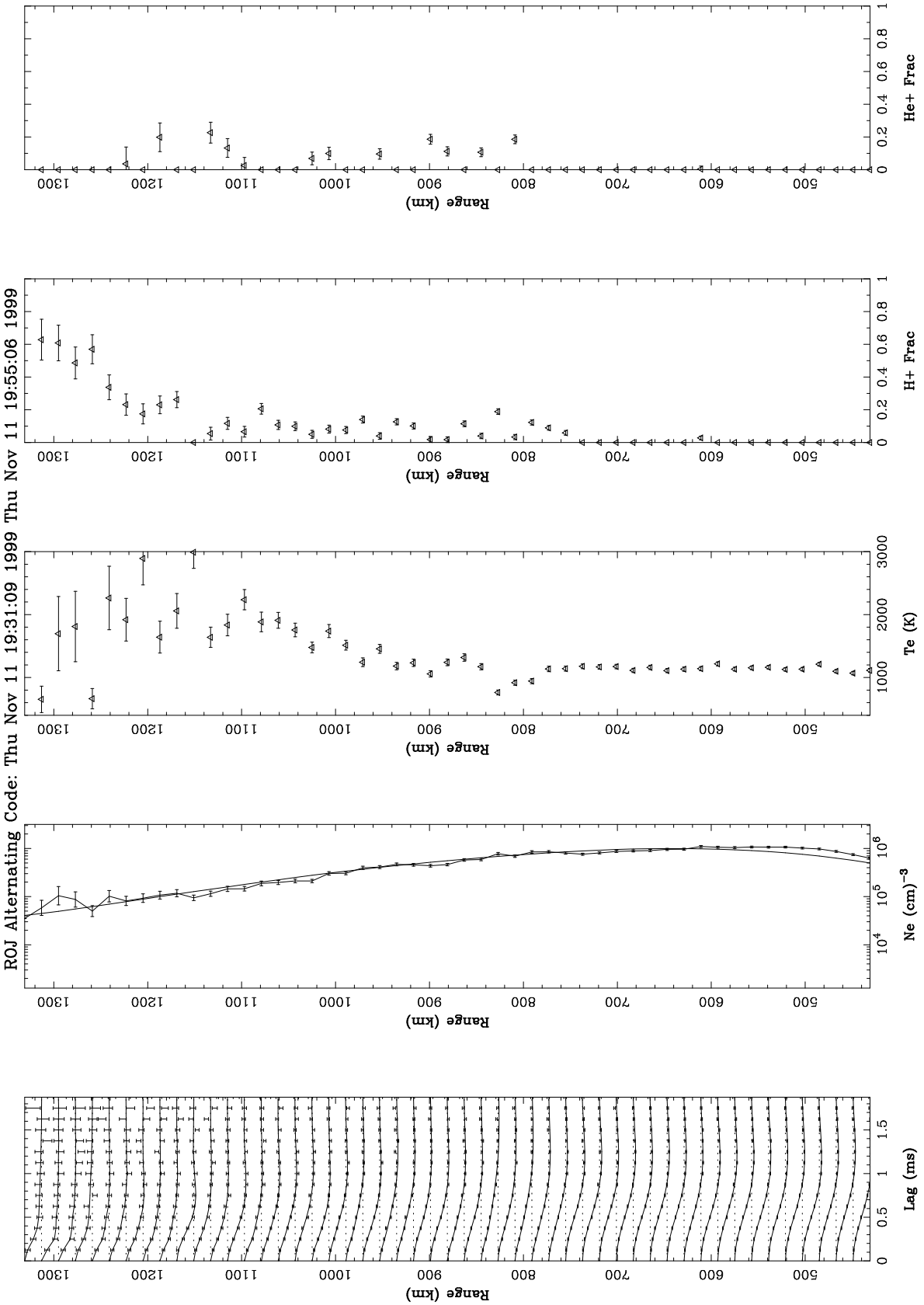

Figure 4. Alternating code incoherent scatter data for November 11, 1999. Two transmitters were used for this experiment. Residual satellite contamination is evident at altitudes close to 820 km.

the National Science Foundation through cooperative agreements ATM-9022717 and ATM-9408441 to Cornell University and by NSF grant ATM-9711236 to Clemson University. The Jicamarca Radio Observatory is operated by the Geophysical Institute of Perú, Ministry of Education, with support from the NSF cooperative agreements just mentioned. The help of the staff, particularly R. F. Woodman, L. Condori, and J. L. Chau, was much appreciated.

References

Bevington, P. R., *Data Reduction and Error Analysis for the Physical Sciences*, McGraw-Hill, New York, 1969.

Farley, D. T., Incoherent scatter correlation function measurements, *Radio Sci.*, *4*, 935, 1969.

Farley, D. T., Multiple-pulse incoherent-scatter correlation function measurements, *Radio Sci.*, *7*, 661, 1972.

Farley, D. T., Early incoherent scatter observations at Jicamarca, *J. Atmos. Terr. Phys.*, *53*, 665, 1991.

Holt, J. M., D. A. Rhoda, D. Tetenbaum, and A. P. van Eyken, Optimal analysis of incoherent scatter radar data, *Radio Sci.*, *27*, 435, 1992.

Huuskonen, A., and M. S. Lehtinen, The accuracy of incoherent scatter measurements: Error estimates valid for high signal levels, *J. Atmos. Terr. Phys.*, *58*, 453, 1996.

Kudeki, E., S. Bhattacharyya, and R. F. Woodman, A new approach in incoherent scatter *F* region E x B drift measurements at Jicamarca, *J. Geophys. Res.*, *104*, 28,145, 1999.

Lehtinen, M. S., Statistical theory of incoherent scatter radar measurements, *Tech. Rep. 86/45*, Eur. Incoherent Scatter Sci. Assoc., Kiruna, Sweden, 1986.

Lehtinen, M. S., and I. Häggström, A new modulation principle for incoherent scatter measurements, *Radio Sci.*, *22*, 625, 1987.

Lehtinen, M. S., and A. Huuskonen, General incoherent scatter analysis and GUISDAP, *J. Atmos. Terr. Phys.*, *58*, 435, 1996.

Lehtinen, M. S., A. Huuskonen, and M. Markkanen, Randomization of alternating codes: Improving incoherent scatter measurements by reducing correlations of gated correlation function estimates, *Radio Sci.*, *32*, 2271, 1997.

Markkanen, M., and T. Nygrén, Long alternating codes, 2, Practical search method, *Radio Sci.*, *32*, 9, 1997.

Nygrén, T., and M. Markkanen, Long alternating codes, 1, Search by playing dominoes, *Radio Sci.*, *32*, 1, 1997.

Pingree, J. E., Incoherent scatter measurements and inferred energy fluxes in the equarotial F-region ionosphere, Ph.D. thesis, Cornell Univ., Ithaca, N. Y., 1990.

Sulzer, M. P., A radar technique for high range resolution incoherent scatter autocorrelation function measurements utilizing the full average power of klystron radars, *Radio Sci.*, *21*, 1035, 1986.

Sulzer, M. P., A new type of alternating code for incoherent scatter measurements, *Radio Sci.*, *28*, 995, 1993.

Sulzer, M. P., and S. Gonzalez, The effect of electron coulomb collisions on the incoherent scatter spectrum in the *F* region at Jicamarca, *J. Geophys. Res.*, *104*, 22,535, 1999.

D. L. Hysell, Department of Physics and Astronomy, 205 Kinard Laboratory, Clemson University, Clemson, SC 29634. (dhysell@clemson.edu)

Received April 10, 2000; revised July 20, 2000; accepted August 7, 2000.