POLAR MESOSPHERE SUMMER RADAR ECHOES: OBSERVATIONS AND CURRENT THEORIES

John Y. N. Cho
Michael C. Kelley
School of Electrical Engineering
Cornell University
Ithaca, New York

Abstract. The remarkably strong radar echoes from the summer polar mesosphere have been an enigma to atmospheric and radar scientists since their discovery more than a decade ago. Since then, more sophisticated radar experiments and in situ rocket measurements have shed some light on the underlying physics and chemistry, and theories have been formulated to explain the generation of the intense radar backscatter and the remarkable physical conditions associated with it. First, we review the key observations and examine the proposed theories. We then evaluate the progress that has been made in understanding this phenomenon and explore its connection to global change, to the newly recognized material referred to as a dusty plasma, and to the highest clouds in the Earth's atmosphere. Finally, we end with suggestions for future research.

INTRODUCTION

The mesosphere, an atmospheric region of decreasing temperature with increasing height, is a particularly difficult region of the atmosphere to observe. Located between about 50 and 90 km in altitude, above the stratosphere and below the thermosphere, it is much too high for airplanes and balloons but too low for orbital satellites to pass through; the only alternative for in situ sampling has been the launching of the occasional rocket. Active remote sensing from the ground has only been possible since the development of powerful radars and, more recently, lidars. General public awareness of the mere existence of the mesosphere is low even today. Because of the historical dearth of interest and information in this layer of our atmosphere, some pundits have dubbed it the "ignorosphere."

The Polar Summer Mesosphere

The polar summer mesopause has long been somewhat of an exception. (The mesopause is the upper boundary of the mesosphere where the temperature reaches a minimum before increasing with height in the thermosphere.) Ever since the summer of 1885, two years after the cataclysmic eruption of Krakatoa, gazers of the twilight sky in the Arctic have been treated to occasional summer displays of "shining night clouds" that were, from the beginning, recognized as being different from other clouds [Backhouse, 1885]. In 1887, photographic triangulation determined the height of the clouds to be 82 km [Jesse, 1887]. All of a sudden, evidence concerning the thickness of the atmosphere, which had previously extended only up to nacreous clouds at about 30 km, more than doubled in height. Furthermore, the so-called noctilucent cloud (NLC) revealed wave and billow structures which were presumably representative of the motions of the ambient air. Interesting dynamics were going on way up there which scientists now had a way of watching. For a recent review of NLCs, we refer the reader to Thomas [1991] and, for a more complete treatise, to the book by Gadsden and Schröder [1989].

Rocket grenades launched during the International Geophysical Year (IGY) of 1957–1958 revealed another peculiarity of the polar mesopause: it was much colder in the summer than in the winter [Stroud et al., 1959]. This observation supported speculation that the NLC were composed of ice that formed in extremely low temperatures. Subsequent studies have confirmed that summer mesopause temperatures are the lowest found in our atmosphere (temperatures below 100 K have recently been reported by Schmidlin [1992]).

Nearly a century after the first sighting of NLCs, another surprise awaited students of the mesosphere. Observations with a newly installed atmospheric radar in Poker Flat, Alaska, showed that the polar summer mesosphere was incredibly adept at scattering radar
waves [Ecklund and Balsley, 1981]. The enormous radar reflectivities could not be explained by any existing radar scattering theories. The phenomenon was so remarkable that it has earned an acronym, PMSE, which stands for polar mesosphere summer echoes [Röttger et al., 1988].

So three obvious questions regarding the polar summer mesosphere are, (1) Why are clouds there? (2) Why is it so cold? (3) Why are radio waves scattered so strongly there?

There is a wide consensus among researchers that the clouds form precisely because of the coldness, i.e., NLCs are composed of ice particles that have condensed out of the thin air (with the help of nucleation sites such as meteoroid dust and cluster ions). The extreme cold is necessary because the water-mixing ratio is expected to be very low at those altitudes. In fact, the lack of NLC sightings in the historical record before 1885, despite the presence of skilled observers of twilight phenomena in the Arctic, suggests that they were optically too thin before that time to be visible to the naked eye. Wegener [1912] surmised that the eruption of Krakatoa in 1883 injected enough water vapor into the stratosphere for NLCs to eventually become visible. The 2-year delay is consistent with the transport of material through the stratosphere and up into the mesosphere [Schröder, 1985].

More recently, Thomas et al. [1989] have proposed that the anthropogenic increase in methane gas (the oxidation of which in the stratosphere is an important source of water in the middle atmosphere) in the industrial era was responsible for the steady increase in NLC occurrence during the last hundred years [Gadsden, 1985]. On the other hand, Gadsden [1990] has countered that a decrease in the mesopause temperature itself, possibly due to anthropogenic changes in the atmosphere, would have the same effect. We point out that Roble and Dickinson [1989] have predicted a cooling of the mesopause with an increase in methane and carbon dioxide using their global upper atmosphere model. Therefore it is very likely that anthropogenic effects on both the water-mixing ratio and the temperature are working together to increase the cloudiness in the summer mesopause.

The theory of why the summer mesopause is so cold has been developed by Lindzen [1981] and modeled by Holton [1983]. The short version of the story is that the global circulation pattern in the mesosphere is one of summer to winter meridional flow, with upward motion at the summer pole cooling adiabatically and downward flow at the winter pole warming compressionally. In effect, the summer mesosphere is being cooled by a global refrigerator which counteracts the heating of the ever-present summer Sun. The refrigeration pump is the breaking of gravity waves that transfer momentum to the mean flow, spinning up the atmosphere at the summer pole and spinning it down at the winter pole; conservation of momentum and mass completes the circulation cell. The sense of flow that fixes the circuit to be a refrigerator rather than a heater is determined by the fact that the breaking waves have phase velocities that are predominantly eastward in the summer and westward in the winter; these are, in turn, determined by the seasonal stratospheric zonal winds which preferentially absorb gravity waves with matching phase velocities, creating the asymmetry. This theory is well accepted among researchers, but the few measurements that have been made of the mean vertical velocities in the polar mesosphere conflict with the predicted values in sign and magnitude [Balsley and Riddle, 1984; Meek and Manson, 1989].

The third question regarding the polar summer mesosphere, i.e., why are radar waves scattered so much from there, does not have an established answer as the first two do and is the subject of this review paper. The resolution of this scattering problem has, in turn, resolved the conflict between the measured and predicted mean vertical flow in the mesosphere.

The MST Radar

Most people are aware that radar was developed during World War II as a military tool for detecting and estimating the range of airborne targets. What is less known is that the basic principle of using radio wave reflection to determine the distance of the reflector was invented much earlier by Sir Edward Appleton to determine the height of the ionized region in the upper atmosphere [Appleton and Barnett, 1926], one of the key contributions which led to Appleton’s Nobel award in 1947. Thus from the beginning, radars have been intimately connected to observations of the atmosphere.

Most people are also aware of the current use of radars by meteorologists to map out regions of cloud cover and precipitation. Most, however, do not know that a handful of extremely powerful radars can routinely take measurements of the ionosphere by detecting the backscatter from inhomogeneities in the radio refractive index due to the thermal fluctuation of free electrons, a surprising capability known as incoherent (or Thomson) scatter radar which was first predicted by Gordon [1958] and realized by Bowles [1958].

But Woodman and Guillén [1974], using one of these powerful radars built to study the ionospheric plasmas, the Jicamarca radar in Peru, showed that they could also detect backscattered signal from virtually the entire atmospheric height range even without the presence of ionization, as long as there was enough inhomogeneity in the radio refractive index. This type of radar scatter is called coherent scatter, and the class of radars capable of making measurements throughout this atmospheric range is called the mesosphere-stratosphere-troposphere (MST) radar. The radio refractive index is dependent on bound electrons in water vapor in the lower troposphere, bound electrons in dry air in the upper troposphere.
and stratosphere, and free electrons in the mesosphere. Variations in the refractive index are generated by turbulence and stratification of the air, which introduce inhomogeneities in the electron distribution.

Because the radio wave is almost entirely scattered by refractive index structures with a dimension equal to the radar half-wavelength (known as the Bragg scatter condition), most MST radars operate in the VHF (very high frequency) band. This choice best matches the radar Bragg scale to the inertial subrange of clear air turbulence (the length scales at which turbulence can exist) which is the most important contributor to the creation of refractive index fluctuations in the middle atmosphere. Thus UHF (ultra high frequency) radars with their shorter wavelengths are not suitable for MST use because their Bragg scales fall in the viscous subrange of turbulence (the length scales at which the air fluctuations are dissipated by viscosity) in the mesosphere and stratosphere. However, as the kinematic viscosity increases with height, in the upper mesosphere even a VHF radar operates in the viscous subrange and therefore is not expected to observe strong radar echoes.

Therefore one of the most surprising discoveries that researchers have made using MST radars is PMSE. The nearly continuous set of observations made between February 1979 and June 1985 by the 50-MHz Poker Flat radar in Alaska first revealed the nature of polar mesospheric echoes and showed that the summer echoes were remarkably different from both wintertime echoes and echoes detected at other latitudes [Ecklund and Balsley, 1981]. The principal characteristic that has puzzled researchers to this date is the enormous radar backscatter cross section exhibited by this region of the atmosphere. Mesospheric echo powers from nonsummer seasons and nonpolar latitudes are orders of magnitude smaller than the polar summer case and can be readily explained by electron density inhomogeneities created by turbulence, which in turn is generated by breaking gravity waves [Balsley et al., 1983; Røyrvik and Smith, 1984]. The same explanation cannot be directly applied to PMSE because at the altitudes at which they occur, the viscous cutoff scale of the neutral air turbulence is much larger than the radar scattering length, which means that the turbulent energy would have been dissipated by viscosity without creating significant structures at the radar scattering scales. The result is that only the huge Jicamarca radar in Peru is capable of detecting 50-MHz echoes at, say, 85 km. Clearly, there must be something extraordinary happening in the summer polar mesosphere that drastically enhances the radar scattering process.

Because of their high frequency of occurrence and large backscattered power, PMSE allow us to observe the dynamics of the polar summer mesosphere even with relatively small radars such as the Max Planck Institute’s sounding system (SOUSY) and the Cornell University portable radar interferometer (CUPRI). Many papers have been published concerning the observation of tidal modes, gravity waves, momentum flux, and turbulence [e.g., Balsley and Carter, 1982; Carter and Balsley, 1982; Balsley and Garelo, 1985; Fritts, 1988; Kelley and Ulwick, 1988; Ulwick et al., 1988; Watkins et al., 1988; Reid et al., 1988; Rishbeth et al., 1988; Röttger et al., 1988, 1990a; Rüster et al., 1988; Williams et al., 1989; Fritts et al., 1990; Inhester et al., 1990; Kelley et al., 1990; Rüster and Reid, 1990]. Since the gravity waves are often in the large-amplitude saturation regime, they provide a rich laboratory for the study of complex, nonlinear behavior. These topics will not be covered in detail in this paper.

PMSE research is currently very active, and there are many puzzles left unsolved. Therefore this review does not purport to be an authoritative explanation of PMSE nor even a comprehensive evaluation of all the work that has been done. Rather, it is meant to be a middle-of-the-term review that will bring together the different aspects of PMSE research and evaluate where we have made progress and where more work needs to be done. We hope it will also stimulate the interest of atmospheric and radar scientists unfamiliar with the subject and encourage their future contributions. With such thoughts in mind we have tried to keep the length relatively brief and free of equations. To place this review in the larger context of middle atmosphere radar studies, we refer the reader to an excellent article by Röttger [1993]. The organization of the paper is as follows: A summary of the existing data is provided which is followed by a discussion of the theories which have arisen to explain these data. A summary and discussion will follow, then suggestions for future research concerning the unanswered questions will conclude the paper.

OBSERVED CHARACTERISTICS OF PMSE

The only long-term observational data base available for PMSE is still the original Poker Flat set, which is available in the CEDAR (coupling, energetics, and dynamics of atmospheric regions) data base (for information, contact W. Emery, University Corporation for Atmospheric Research, Boulder, Colorado, e-mail emery@ncar.ucar.edu). It is the best source for climatological information. For higher resolution (spatially, temporally, and spectrally), data from radars that have been operated on campaign bases are available. Among those radars, a distinction should be made between VHF studies (Aberystwyth, CUPRI, European incoherent scatter (EISCAT) 224-MHz, Poker Flat, SOUSY) and UHF studies (EISCAT 933-MHz, Sondrestrom), since at the UHF frequencies PMSE are much rarer and weaker. In the mesosphere the UHF radars are used primarily for incoherent scatter observations, while the VHF radars, owing to their
longer wavelength, can also double as coherent scatter detectors. Interferometry data, which yield information about discrete scatterers within the radar volume, are available for some of the EISCAT VHF and CU-PRI observations, and three-dimensional velocities (using multibeam configurations) were measured by the Poker Flat and SOUSY radars.

Climatology

Both the long-term Poker Flat data and the shorter sets from other radars show that important features such as height of occurrence appear to remain stable from year to year. Figure 1 shows PMSE (as marked by echoes at the higher altitudes) beginning and ending fairly abruptly during the middle of the year. Balsley et al. [1983] showed that the PMSE season corresponds well to the polar summer season. The season of noctilucent cloud occurrence [Gadsden, 1982] is also very similar to the PMSE season. We know that an extremely low temperature is one of the necessary conditions for the formation of NLCs. It seems likely that PMSE also require low temperatures.

PMSE have been observed from as far north as Tromsø, Norway (69°35'N) [Hoppe et al., 1988], and as far south as the Harz mountains of Germany (52°N) [Reid et al., 1989] and Aberystwyth, Wales (52.4°N) [Thomas et al., 1992]. (Echoes of the latter type are technically MSE because they are no longer in the polar region.) Radars operating further south have not observed PMSE-like phenomena. Observations by a 2.9-MHz radar at Scott Base, Antarctica (78°S), have not exhibited the peculiar echo characteristics [Fraser and Khan, 1990], but neither have data from the 2.78-MHz partial reflection experiment (PRE) radar in Tromsø [Hoppe et al., 1990]. The characteristics peculiar to PMSE seem to occur only at radar frequencies of VHF or higher. Recently, a team of scientists from the United States and Peru installed a VHF radar in Antarctica and were astounded at the absence of PMSE (B. B. Balsley et al., Southern-hemisphere PMSE: Where are they?, submitted to Geophysical Research Letters, 1993, hereinafter referred to as Balsley et al., 1993).

Low temperature is likely the key factor in the latitudinal distribution of PMSE, since the mesopause temperature rises as one moves from the summer toward the winter pole. NLCs also are rarely observed below about 50° in latitude [Fogle, 1966]. Note that optical observation of these clouds is not possible above ~70° latitude. This is because the clouds are so tenuous that the Sun must dip at least 6° below the horizon such that the sky has darkened considerably but the clouds are still lit from below. It follows then that even within the latitudes of optimal observational conditions, noctilucent clouds can only be seen during a small time window during twilight.

Figure 1. The altitude versus season distribution of mesospheric radar echoes observed by the Poker Flat, Alaska, system [Ecklund and Balsley, 1981]. Hatching indicates no data.
Since 1972 [Donahue et al., 1972], satellites have been observing thin scattering layers in the polar summer mesopause. Dubbed polar mesospheric clouds (PMCs), they form patches that extend over the entire summer polar cap. Extensive measurements made by the Solar Mesosphere Explorer (SME) satellite between 1982 and 1986 [Olivero and Thomas, 1986; Thomas and Olivero, 1989] have generally supported the contention that PMCs and NLCs are manifestations of the same phenomenon. Some differences do exist, however. A comparison of Arctic PMCs and NLCs showed that the southern border of the latter was typically located about 10° below the southern border of the former [Gadsden and Schröder, 1989]. Also, the average cloud particle radius has been measured to be less than 0.06 μm for PMCs [Thomas and McKay, 1985], whereas ground-based observations of NLCs have yielded values as high as 0.3 μm [Gadsden, 1975]. These apparent differences suggest an interpretation of PMCs as a "nursery" of NLCs [Gadsden and Schröder, 1989]. NLCs may consist of large particles that occasionally develop from the embryonic PMC aerosols and are blown equatorward by the prevailing winds. Whatever the precise situation is, in this paper we will simply refer to the clouds as NLCs in keeping with the more historical nomenclature.

Although it is tempting to postulate a link between the clouds and the radar echoes, a one-to-one correspondence is immediately ruled out because NLCs occur during only a small fraction of the time that PMSE are observed. Taylor et al. [1989] have even observed a case where NLC were present without PMSE, and the comparisons between the EISCAT VHF radar, CUPRI, and ground-based visual observations of NLC during the summer of 1991 showed no correlation between PMSE and NLC occurrence (S. Kirkwood et al., A comparison of PMSE and other ground-based observations during the NLC-91 campaign, submitted to Journal of Atmospheric and Terrestrial Physics, 1993). On the other hand, Jensen et al. [1988] have measured a weak correlation between Poker Flat PMSE and satellite UV observations, and in one instance rocket-borne instruments detected an NLC layer near the CUPRI radar volume which also registered PMSE simultaneously (U. Wfilchli et al., First height comparisons of noctilucent clouds and simultaneous PMSE, submitted to Geophysical Research Letters, 1993). To explain these ambiguous results it may be argued that subvisible particles may be present even when clouds are not seen from the ground. Furthermore, from the ground, NLCs are only observable just after sunset and before dawn, when the solar ionization rate is low and fewer electrons are available for radar scattering. A theory of how aerosols might influence the radar echoes will be discussed in the theory section.

Figure 2. Time-averaged altitude profiles of signal-to-noise ratio (in decibels) for typical summer and winter periods [Ecklund and Balsley, 1981].

Morphology and Temporal Variability

Figure 2 shows averaged signal-to-noise profiles for PMSE versus nonsummer echoes for the Poker Flat radar. Note that PMSE occur in a well-defined layer between 80 and 93 km with a peak around 86 km. These statistics are closely matched by the SOUSY 53.5-MHz radar data taken from Andøya, Norway [Czechowsky et al., 1989], by the EISCAT 224-MHz radar data taken from Tromsø, Norway [Hoppe et al., 1990], and by the CUPRI 46.9-MHz radar data taken from Esrange, Sweden [Cho, 1993].

One point to keep in mind while viewing nonscanning radar data is that apparent changes in the scattering structure are due to both the time evolution of the structure itself and advection of the structure through the radar beam by background winds. Range-time-intensity plots must be interpreted with this in mind. The top panel of Plate 1 is an example of such a plot produced from CUPRI data. Note that individual scattering structures can be thinner than the height resolution here of 300 m (perhaps even less than 100 m according to Franke et al. [1992]) and that multiple layers can occur.

Echo power fluctuations within a minute or less are usually unrelated vertically and therefore not caused by the precipitation of energetic particles which create vertically extended regions of enhanced ionization.
Plate 1. Plots of signal-to-noise ratio (top), vertical velocity (middle), and Doppler spectral width (bottom) versus altitude and time as observed by the CUPRI [Cho, 1993]. The height resolution is 300 m and the time resolution is 34 s.

[Röttger et al., 1988]. Both Luhmann et al. [1983] and Czechowsky et al. [1989] reported a low correlation between PMSE power and the ambient electron density level, whereas they showed that the winter echoes are highly dependent on precipitation events. The short-term variability in the summer is thought to be the result of "blobs" or "crinkled" layers being advected horizontally through the radar volume. In general, ionization enhancement due to precipitation can further "illuminate" already existing PMSE layers such that the overall signal increases. However, the baseline solar ionization level during the summer (the Sun is continuously shining on the mesosphere) seems to provide enough electron density around the clock such that observation of PMSE is not dependent on extra electrons produced by precipitation events (at least for the more sensitive of the VHF radars).

Early theories concerning the production of radar scattering structures in the summer mesosphere centered on the instability of atmospheric tidal modes and gravity waves. Enhanced radar backscatter was thought to be caused by regions of intense turbulence produced by regions of instability.

Balsley et al. [1983] proposed that the shear instability of tidal modes was the dominant source of turbulence. Semidiurnal variations in the echo power intensity consistent with this hypothesis were observed by Czechowsky et al. [1989], but they also noted that in their data set, regions of maximum wind shear were not always correlated with areas of largest echo return (Figure 3) which ran counter to the shear-instability theory. Diurnal variations in the echo intensity were also observed in the long-term average of Balsley et al. [1983], which showed a significant dip in power around 2200 LT.

Supersaturation of gravity waves, an enhancement of small-scale waves due to the compression of vertical scales as the waves enter a region of higher stability above the mesopause, was invoked by VanZandt and Fritts [1989] as a generation mechanism for strong turbulence. Fritts [1988] showed a case where the regions of maximum radar backscatter matched the progression of the most unstable phase of a 7-hour wave inferred from the temperature and velocity data (Figure 4). Now, these various theories may well be correct in explaining the generation and modulation of turbulent mixing in the polar summer mesosphere. However, when viewed in the context of an explanation for the raw intensity of the echo, they are totally inadequate.

There are, in fact, serious defects with a classic turbulent radar scatter explanation of PMSE. For example, Ulwick et al. [1988] quoted values for the turbulent dissipation rate and electron density during the Structure and Atmospheric Turbulence Environment (STATE) rocket/radar experiments at Poker Flat
Figure 3. Plots of wind shear, echo power, and Doppler spectral width versus height and time as observed by the SOUSY radar [Czechowsky et al., 1989].

nearly identical to those measured in Peru during the Condor rocket campaign [Rørvik and Smith, 1984] but with a radar cross section that was 4 orders of magnitude higher. Similarly, the turbulence intensities inferred from the Doppler spectral widths measured at EISCAT were much too low to produce the extremely large radar signals according to the classical theory [Röttger and La Hoz, 1990]. Furthermore, as we shall see in the section on aspect sensitivity, there are other reasons to believe that the turbulence, per se, is not the most important element in producing the intensity or even the occurrence of PMSE. Possible resolutions to this problem will be discussed in the theory section.

Overall, regions of strong echoes more often appear to drift downward (with velocities of the order of 1 m/s) rather than upward [Balsley et al., 1983]. Since the phase velocities of gravity waves and tides have downward components, the observations suggest that the radar-scattering regions correspond to a certain phase of the waves, such as maximum instability (turbulence generation), maximum upward velocity, or minimum temperature. Correlation of upward velocity with echo power has been observed by Williams et al. [1989] during one observation period, but such a correspondence does not always hold.

Evidence of short-period gravity waves can also be observed in the Doppler scale of the center panel of Plate 1. Röttger et al. [1990a] have noted frequent examples of steepening and tilting of Doppler velocity wave structures which imply a nonlinear transfer of energy from the fundamental wave frequency to higher harmonics. The vertical Doppler spectrograms of CUPRI data in Figure 5 show several examples of highly distorted sinusoidal structures propagating or advecting through the radar beam. The amplitudes of gravity waves that travel up into this altitude region can be-
Simultaneous observations by the CUPRI 46.9-MHz and the EISCAT 224-MHz radars revealed that the evolution and spatial structures of PMSE at those frequencies were very similar (see Figure 6). Thus the same radar-scattering mechanism is likely to be producing PMSE at those frequencies.

Figure 7 shows a comparison of CUPRI and EISCAT 933-MHz data taken simultaneously. At the higher frequency, the EISCAT radar is normally expected to detect only incoherent scatter with the echo power dependent almost solely on the electron density. In the left-hand panels the normal incoherent scatter dominates and the EISCAT profile can be interpreted as electron density, while the CUPRI profile shows an obvious PMSE layer. Note the dip in the EISCAT signal-to-noise ratio (SNR) profile at 85 km that corresponds to the peak in the CUPRI SNR. Such “bite-outs” in electron density have also been measured by rocket probes (see the “comparison with rocket measurements” subsection below). In the panels on the right side of Figure 7, a peak in the EISCAT SNR profile can be seen in the altitude region of the CUPRI PMSE. This layer observed at UHF had a Doppler spectral width much narrower than the incoherent scatter spectra in the other range gates [Röttger et al., 1990b], indicating that the echo power enhancement was not simply due to a thin slab of extra electrons (which, in any case, would be almost impossible to explain with known electron generation mechanisms), but a manifestation of PMSE which may or may not have been produced by the mechanism that caused the VHF PMSE.

Spectral Width

One of the first things to be gleaned from the PMSE Doppler data was that the spectral widths were too narrow for classic incoherent scatter at both VHF [Röttger et al., 1988] and UHF [Röttger et al., 1990b]. Thus PMSE were shown to be coherent echoes, which have inherently narrower Doppler spectra, from this perspective as well as from echo strength considerations.

The spectral widths from coherent scatter radars can yield information regarding the scatter/reflection mechanism. In short, radar scatter from a turbulent medium has a spectrum with a width that is correlated with the turbulence energy. In other words, the more turbulent the medium is, the greater the spread in Doppler velocities. (This is a simplified explanation, since other spectral broadening factors must be taken into account before the extraction of turbulence information from the spectral width [Hocking, 1985].) Fresnel reflection or scatter generally produces a narrower spectrum than turbulent scatter and points to a medium that is coherent in the plane perpendicular to the radar beam. Figure 8 is a plot of coherence time (inverse of spectral width) versus echo power for the EISCAT 224-MHz radar. As the echo power grows,
we see little correlation with the coherence time. In fact, at the upper range gates the coherence seems to decrease as the echo strength increases. This is a fairly typical example at 224 MHz. Examination of data from a different radar, such as the 46.9-MHz CUPRI, also shows that backscattered power can range over many decibels while maintaining a rather constant spectral width, i.e., a constant apparent turbulence intensity.

**Aspect Sensitivity**

Another piece of evidence for the complexity of the echoing process is the dependence of the echo power on the pointing direction of the radar beam. Czechowsky et al. [1988] noted that the backscattered signal level at 50 MHz decreased for antenna beams pointed away from the vertical (Figure 9). If the classic homogeneous isotropic turbulence of Kolmogorov [1941] were responsible for the radar echoes, one would expect the scattering structure to be isotropic, and hence the scattered power should not depend on the direction of the radar beam. Because the aspect sensitivity is centered around the vertical, i.e., gravitational, axis and since the echo structures are horizontally layered, one can infer that stratification plays an important role in the formation of PMSE layers and their scattering properties.

**Comparison With Rocket Measurements**

There have been three major radar/rocket campaigns in the polar summer mesosphere: STATE in 1983, the Middle Atmosphere Cooperation/Summer in Northern Europe (MAC/SINE) in 1987, and the Noctilucent Cloud-91 (NLC-91) in 1991. Papers from the first campaign were printed in the June 20, 1988, issue of the Journal of Geophysical Research; articles from the second were published in a special issue (October-November 1990) of Journal of Atmospheric and Terrestrial Physics; and reports from the third will be collected in future issues of Geophysical Research Letters and Journal of Atmospheric and Terrestrial Physics.

The first rocket probe measurements of electron density simultaneous with VHF polar radar data [Ul-
wick et al., 1988] revealed two important clues to the PMSE puzzle: (1) the electron density was, in fact, structuring at length scales well below the viscous cutoff of the neutral gas [Ulwick et al., 1988], and (2) sharp depletions or bite-outs were often present at the altitudes of the echo layers [Kelley and Ulwick, 1988].

Result 1 meant that radar reflectivities calculated directly from the rocket-derived electron density fluctuation power spectra, using isotropic turbulent scatter theory, agreed reasonably well with the actual radar results (Figure 10). Thus the suspicion was shifted from the radar backscatter theory to the assumption that the electrons were behaving as perfectly passive tracers of the neutral gas, a theory that indeed yielded excellent quantitative agreement with the much weaker equatorial mesosphere echo strength [Røysvik and Smith, 1984]. This assumption had seemed reasonable in the past, since the number of electrons is about $10^{-10}$ of the neutrals. Clearly, some other factor is keeping the electrons from simply mimicking the structure of the neutral gas. Solutions to this dilemma will be discussed in the theory section.

Result 2 suggested that sharp gradients in the electron density are an important factor in producing the radar echo layers and may help to explain their aspect sensitivity. Once again, the mystery centers on the mechanisms that could generate and maintain such sharp bite-outs in the electron density. These abrupt depletions may signal the presence of a layer of aerosols that scavenge the ambient electrons. We will show later that such charged aerosols are also potentially very important in the generation of enhanced radar echoes.

During the NLC-91 campaign, a rocket equipped with both a neutral density sensor and plasma probes flown through a two-layer PMSE showed almost no neutral turbulence in the lower layer and classic turbulence in the upper layer; the plasma had nonturbulent fluctuations below and turbulent fluctuations above [Lübken et al., 1993; J. C. Ulwick et al., Evidence for two different structuring and scattering mechanisms and the associated role of aerosols in the polar summer mesosphere, submitted to Geophysical Research Letters, 1993, hereinafter referred to as Ulwick et al., 1993]. Corresponding to these rocket observations, both the CUPRI and the EISCAT VHF radar showed Doppler spectra that were much wider in the upper layer than in the lower layer [Hoppe et al., 1993; Cho et al., 1993]. Furthermore, the echoes in the upper region were isotropic, whereas in the lower region they were aspect sensitive [Cho et al., 1993]. Thus at least in this particular instance, there were two distinct types of PMSE: induced by neutral turbulence in the upper layer and by something else in the lower layer.

**Mean Vertical Velocity**

Finally we come to one last piece of strange behavior as observed by the radars: the mean vertical velocity measured during the summer is substantially downward ($-20$–$30$ cm s$^{-1}$) [Balsley and Riddle, 1984; Meek and Manson, 1989]. This result is in opposition to the theoretical requirement that the summer polar mesosphere be flowing upward in order to cool it down to the observed temperatures that are far below the radiative equilibrium values [Lindzen, 1981].
Figure 7. Signal-to-noise ratio (SNR) versus height for (left) the EISCAT 933-MHz radar and (middle) the CUPRI 46.9-MHz radar, and (right) the vertical gradient of the EISCAT-inferred electron density in units of m^{-4} versus height [Röttger et al., 1990b]. The small dots are averaged over 1 min, and the large dots are averaged for the indicated time per panel.

shows a multiyear composite of monthly vertical velocity means observed by the Poker Flat radar [Hall et al., 1992]. Note the significant downward values in the summer in the upper mesosphere. Theories predict a circulation pattern in the summer of an upward velocity of ~1 cm s^{-1}, an equatorward meridional flow, and a westward zonal mean flow. The latter two conditions were observed by the radar, but the upward flow clearly was not measured. In fact, the winter values (which should be downward) appear to be slightly in the opposite direction to the theoretical prediction as well. Shorter data sets taken by the CUPRI in northern Scandinavia have also yielded a mean downward velocity in the summer [Hall, 1991; Cho, 1993].

A second-order, compressional gravity wave effect called the Stokes drift was invoked by Coy et al. [1986] to reconcile the measurements with theory. The Stokes drift is the difference between averages made in Eulerian and Lagrangian reference frames in the presence of significant wave fluctuations. A radar, under certain assumptions, measures an Eulerian velocity, whereas the theoretical circulation is a Lagrangian motion. However, this mechanism is not seasonally dependent and cannot account for the fact that large
downward velocities are not observed in the winter as well as the summer. Moreover, Hall et al. [1992] showed that the Stokes drift would be about an order of magnitude smaller than the measured mean velocity for a realistic spectrum of gravity waves rather than the monochromatic wave used by Coy et al. [1986].

Hall [1991] also considered the effect of tilted scattering layers advected by horizontal winds projecting an apparent vertical velocity onto the radar Doppler measurement. He rejected this idea after failing to find evidence for a correlation between the preferred gravity wave horizontal propagation direction and the mean horizontal velocity vector in the Poker Flat data.

After rejecting a number of other effects as well, Hall et al. [1992] proposed that the radars have been measuring the fall velocity of charged ice aerosols that may also be the key to the generation of PMSE (see theory section for explanation). This mechanism has the advantage of turning off during the nonsummer seasons when the temperature gets too high for the particles to form, thus matching the observed seasonal behavior. However, it is not clear whether this idea could explain the mean downward flow observed by the Saskatoon MF radar [Meek and Manson, 1989] which, at the longer Bragg scale, may or may not be affected by the charged aerosols.

Stitt and Kudeki [1991] suggested that the distortion of waveforms in large-amplitude gravity waves can cause preferential sampling of certain phases of the velocity field. In general, velocity measurements by MST radars in the presence of waves is currently a topic of active debate. A more quantitative study needs to be conducted to evaluate whether such effects are important in the case of PMSE.

Summary of Observations
We list below a summary of the pertinent points from the observations.

1. In the polar summer mesosphere, the VHF radar cross sections are enormously enhanced relative to those in other seasons and latitudes.
2. Rocket measurements show that in the PMSE layers, the electron density has structures well below the viscous cutoff length scale of the neutral gas.
3. The range of occurrence of PMSE in space and time corresponds well to the cold summer mesopause.

Figure 8. EISCAT VHF scatter plots of coherence time versus echo power [Röttger and La Hoz, 1990]. Each dot is a 30-s average. Units of coherence time are 10^{-2} s.

Figure 9. Plots of backscattered power relative to that received in the vertical beam as a function of zenith angle for three different periods [Czechowsky et al., 1988]. The angle \( \theta \) is the half width inferred from the small-angle data pairs assuming the indicated exponential fall-off function.
In the same region, large ions and aerosols form due to the uniquely low temperatures.

4. Very steep gradients and "bite-outs" in the electron density are often observed in PMSE regions by rocket probes and the EISCAT UHF radar.

5. Semidiurnal periodicities in PMSE strength are apparent, and specific cases have been observed in which regions of maximum echo power corresponded to an unstable phase of tidal modes and long-period gravity waves.

6. Short-period gravity waves are often observed to be steepened, tilted, and sometimes saturated with height.

7. PMSE at 50 MHz are reported to be aspect sensitive with respect to the vertical.

8. At VHF, the Doppler spectral widths are not positively correlated with backscattered power.

9. In at least one instance, two distinct conditions (turbulent and nonturbulent) were observed by rockets that flew through a PMSE double layer.

10. The radar cross sections are highly frequency dependent, but the behaviors at 50 and 224 MHz are very similar and can most likely be explained by the same process. PMSE at 933 MHz and 1.29 GHz are much rarer and weaker but nonetheless much stronger than could be explained by reasonable extrapolations of the VHF echoes.

Observation 2 is the direct explanation for observation 1. However, observation 2 itself is a puzzle that needs an answer, and a large portion of the following theory section will be devoted to possible solutions which revolve around item 3 and may also have a role in explaining 4. Items 5 and 6 suggest that turbulence does exist in the summer mesosphere and affects the generation of PMSE; however, 7 and 8 imply that the turbulence, in general, is not the only, or perhaps even the key, element in producing PMSE. Observation 9 suggests that PMSE may be clearly divided into two categories (turbulent and nonturbulent), and observation 10 hints that PMSE at UHF may be fundamentally different from PMSE at VHF.

**THEORIES OF PMSE**

A radio wave traveling in a vacuum will keep going forever, undisturbed, barring any general relativistic effects. However, a perfect vacuum is rarely encountered in real life. In almost all cases the medium of propagation will be filled with various obstacles that disturb and scatter the wave. In Earth's atmosphere,
radio waves encounter bugs, birds, rain and snow, clouds, airplanes, parachutists, ICBMs, UFOs, meteoroids, etc., all of which can modify the propagation. The atmosphere itself can scatter radio waves when, for a variety of reasons, it has variations in the index of refraction. Radar scientists take advantage of this phenomenon to obtain information about the "target" by analyzing the scattered signal.

In the mesosphere the dominant source of such variations in the radio refractive index is the fluctuation in the free electron density. (Direct Rayleigh scattering from noctilucent cloud particles is much too weak to make a detectable contribution.) We can categorize the electron density fluctuations as follows: (1) the irreducible minimum resulting from the thermal energy and (2) anything above that which is created by various mechanical, electrodynamical, and chemical effects. Type 1 is termed Thomson (incoherent) scatter, and type 2 is referred to as coherent scatter. In both cases the waves are scattered back by inhomogeneities of scale corresponding to half the radar wavelength (known as the Bragg scatter condition). Because of the high strength of the echoes and the narrowness of the Doppler spectra, PMSE must be the result of coherent scatter. However, as was discussed in the observation section, there are some reasons to believe that PMSE observed by UHF radars may be generated by a different mechanism than those detected by VHF radars. We will first examine the theories that have been developed for the VHF data.

**VHF PMSE**

The existence of electron density structures can be thought of as a continuous struggle between the generation mechanisms and the ever present diffusive dissipation (see Figure 12). Diffusion acts preferentially on shorter length scales, so in general it is harder to maintain smaller structures. Hence the central problem of PMSE: what anomaly allows the maintenance of electron density inhomogeneities at smaller scales (i.e., the radar Bragg scales) than are normally possible in the mesosphere? The short answer is that either (1) the generation of structures is enhanced or (2) the electron diffusivity is reduced, or possibly both.

**Enhanced generation mechanisms.** Generation of electron density structures in the mesosphere can be due to the following mechanisms: (1) dynamics of the neutral gas, (2) chemistry, and (3) electrodynamics. Magnetic field effects can be ignored because of the high collision frequencies and because the PMSE structures are not dependent on the magnetic field direction.
Early theories focused on the effects of neutral dynamics on the electrons because there are so few electrons in comparison with the neutral molecules; the electrons were assumed to be passive tracers of the neutral gas. Balsley et al. [1983] attributed the generation mechanism to turbulence produced by shear instability of tidal modes. They did, however, realize that the intensity of typical turbulence is not great enough in the summer mesosphere to account for the observed radar cross sections, a point that was confirmed by later measurements with the EISCAT and SOUSY radars.

Turbulence and stratification are somewhat at odds with each other in terms of the stability criterion, although both can exist simultaneously. In the extremes, intense turbulence exists in an unstable environment and mixes the fluid properties isotropically, while strong stratification is by definition stable and creates an anisotropic situation where the fluid properties are horizontally coherent but may be sharply changing vertically. The radar-scattering types that correspond to these two extreme situations are (1) isotropic turbulent scatter and (2) Fresnel (partial) reflection. Type 1 is statistical whereas type 2 is deterministic, reflecting the nature of the media. Two intermediate cases, anisotropic turbulent scatter and Fresnel scatter (both statistical), are often delineated; Doviak and Zrnić [1984] have formulated a theory that unifies them all under one framework.

As for PMSE, the most favored scenario has been that of weak turbulence coexisting with vertically steep, horizontally coherent structures in the electron density, i.e., a combination of weak turbulent scatter and Fresnel scatter. Such a picture is supported by the following evidence: (1) billows in noctilucent clouds [Witt, 1962] and saturating gravity waves [Fritts, 1988] imply that turbulence exists near the summer mesopause, (2) radar spectral widths indicate mostly weak turbulence [Watkins et al., 1988; Röttger and La Hoz, 1990], (3) radar aspect sensitivity shows isotropy (weak, background turbulence) at large angles and a sharp fall-off in backscattered power with increasing zenith angle at small angles (Fresnel scatter) [Czechowsky et al., 1988], and (4) simultaneous rocket and radar data often show sharp edges in the electron density profile sometimes with and sometimes without fine structure [Ulwick et al., 1988]. Even so, we still need a plausible mechanism that can create coherent structures at the short radar Bragg scales; stratification of the neutral gas alone is not likely to produce such sharp features.

More recently, however, evidence has been accumulating that PMSE can be instigated separately by both isotropic turbulence and Fresnel scatter on different occasions [Thomas et al., 1992] and simultaneously in different layers [Labken et al., 1993; Cho et al., 1993; Ulwick et al., 1993].

Gravity waves can produce local minima in the vertical temperature profile where enhancement of electron recombination, water cluster formation, and nucleation of ice aerosols can take place. The direct effect of the waves on the electron density is large-scale, i.e., of the order of their vertical wavelengths [Sugiyama, 1988], and not likely to cause radar scatter, but Röttger and La Hoz [1990] have suggested that the ice particles while in the updraft phase of a wave may undergo a charge separation process akin to those in tropospheric convective clouds, which can develop electric fields and electron density fluctuations at short scales. In fact, there has been a rocket measurement of a dc electric field inside a noctilucent cloud [Goldberg, 1989]. As we shall see later in this section, the presence of charged aerosols is likely to be a key condition for PMSE occurrence, so the gravity-wave-produced local temperature minima may explain certain behaviors of PMSE such as two layers tracking each other over time. Unfortunately, the theory for aerosol charging in the mesosphere is not well developed, and relevant laboratory measurements are practically nonexistent. Much more work needs to be done in this area.

Another scheme that relies on the interaction of charged aerosols with the ambient air flow was proposed by Havnes et al. [1992]. Assuming that the existence of vortices embedded in a field of falling aerosols yields "holes" inside the vortices into which the aerosols cannot penetrate. A sharp gradient in the aerosol density (and correspondingly the electron density) develops at the edge of each vortex, resulting in partial reflection of radar waves. Havnes et al. [1992] were able to produce reasonable values of reflectivity for VHF radars. However, some questionable assumptions were used in the calculations such that one must be wary of the final results. First of all, the existence of vortices with diameters of the order of 2–4 m is a tenuous proposition, since the viscous cutoff scale for velocity fluctuations in the mesosphere is in the tens of meters. Second, the use of an aerosol size spectrum that extends well above 0.1 μm assumes a noctilucent cloud condition; it is not likely that such large particles are present most of the time. Finally, in the calculation of the radar reflectivity, the physical cross section of the vortex is used, whereas the radar backscatter is likely to come only from very small sections of each vortex wall owing to their high curvatures.

There is one more idea: the presence of dust is known to lower the threshold for shear instability onset in gases. However, the ratio of dust to air mass density must approach an appreciable fraction of unity for this effect to become important [Palaniswamy and Purushotham, 1981]. For the mesopause, this critical ratio is expected to be less than $10^{-5}$ [Havnes et al., 1992], so we can discard this possibility.

**Reduced diffusion mechanisms.** Even if the generation of electron density structures is not especially...
increased, the reduction of electron diffusivity can extend the maintenance of structures to smaller length scales. As an example, we look at isotropic turbulence scatter theory, since a relevant formulation describing the behavior of passive scalars mixed by turbulence has been developed [Batchelor, 1959].

The spectrum of velocity fluctuations in a turbulent medium with high Reynolds numbers was predicted by Kolmogorov [1941] and has been proven to be remarkably accurate in the description of turbulence in the atmosphere and many other fluid media. The basic scenario is as follows: (1) energy from large-scale mean flows and waves are converted to three-dimensional turbulent eddies via instabilities, and (2) vortex stretching mechanisms transfer the eddy kinetic energies to smaller eddies until (3) the eddies become so small that molecular viscosity begins to destroy them. The key postulate is that in the scale range between the large-scale energy input and the small-scale energy dissipation, there exists a range where the kinetic energy spectrum depends only on the energy dissipation rate. Then simply on dimensional grounds, the one-dimensional velocity fluctuation energy spectrum in the cascade region, called the inertial subrange, can be shown to be proportional to $k^{-5/3}$, where $k$ is the wavenumber. (The three-dimensional spectrum, which corresponds to what the radar measures, goes as $k^{-11/3}$. ) There is a clear cutoff of the inertial subrange at a length proportional to the Kolmogorov microscale, $\eta = \nu^{3/4} \varepsilon^{-1/4}$, where $\nu$ is viscosity and $\varepsilon$ is the turbulence dissipation energy, beyond which the spectrum falls off exponentially [Corrsin, 1964]. (Experiments indicate that the constant of proportionality for air density fluctuations is about 7.4 [Hill and Clifford, 1978], that is, the fluctuation spectrum falls off at wavelengths less than $7.4\eta$.)

Herein lies the crux of the problem with classical turbulence scatter theory as applied to PMSE: at the PMSE altitudes the viscous cutoff is in the tens of meters, much longer than the radar Bragg lengths, and thus the radars should not have had significant structures from which to scatter, since they would have been dissipated by viscosity. However, this line of reasoning assumed that the electron density fluctuations, which is what the radar really sees, had the same spectral cutoff as the neutral air density fluctuations, which is only true if the Schmidt number defined by $Sc = \nu/D$, where $D$ is the electron diffusion coefficient, was $\sim 1$. Such a condition holds true in the upper mesosphere under normal circumstances [Hill and Bowhill, 1976; Røyrvik and Smith, 1984]. However, if for some reason the electron diffusivity is drastically reduced, then $Sc$ will become much greater than 1 and an extension beyond the viscous cutoff called the viscous-convective subrange will appear in the fluctuation spectrum of the electron density with a depen-
dence of $k^{-1}$ ($k^{-3}$ for a three-dimensional spectrum) [Batchelor, 1959], a prediction that has been confirmed by experiments [Gibson and Schwartz, 1963].

As a result, structures in the electron density would be maintained to smaller lengths, i.e., the VHF radar Bragg scales. As an aside, we would like to add a cautionary note to the calculation of the viscosity in the cold summer mesopause. The temperatures there may get so low that commonly used expressions, such as Sutherland's formula and empirically determined temperature power laws which are employed to calculate viscosity, can become inaccurate. For example, the viscosity formula used by Banks and Kockarts [1973] for molecular oxygen, while accurate to within 5% of measured values down to 200 K, overestimates laboratory values [Johnston and McCloskey, 1940] by 16% at 130 K.

Kelley and Ulwick [1988] used the preceding ideas in their analysis of the STATE rocket and radar data. Their rocket-measured electron density fluctuation spectrum is contrasted with those from the polar winter mesosphere and the equatorial mesosphere in Figure 13. One can see that the electrons structure at scales smaller than the radar Bragg scale (and the viscous cutoff) only in the polar summer. Also, Hall and Brekke [1988] have inferred larger $S_c$ during the summer than in the winter from the EISCAT VHF incoherent scatter spectral widths. The summer values were not outstandingly large (up to $S_c \sim 5$), but by definition they could not make the measurements in the presence of PMSE, so higher values of $S_c$ could not be expected. To fit the observed PMSE cross sections, Hall [1991] has estimated $S_c$ of 100 to 900 using the rocket and radar data from the MAC/SINE campaign.

Note that a reduction of electron diffusivity also increases the probability of the maintenance of steep vertical gradients in the electron density necessary for Fresnel scatter and reflection. For significant echoes to be produced by a Fresnel-type mechanism, the vertical structure itself must be of the order of the radar Bragg scale [Woodman and Chu, 1989]. Thus independent of whatever mechanism that produces the sharp gradients, it seems likely that reduced electron diffusion is a requirement for preserving the structures at the short Bragg scales.

But what mechanism peculiar to the polar summer mesosphere could be slowing down the electron diffusion? Kelley et al. [1987] suggested that the presence of heavy hydrated ions ($H^+(H_2O)_n$) could "drag" down the electrons via ambipolar electric fields. Such cluster ions only form in the extremely low temperatures of the polar summer mesopause region. Rockets have measured up to $n = 21$ [Björn and Arnold, 1981], and larger clusters have been produced in laboratory simulations [Yang and Castleman, 1991]; Hall [1990] has studied the effect of gravity waves on their formation. Cho et al. [1992a] calculated, however, that proton hydrates with $n < 20$ will not sufficiently reduce diffusivity to raise $S_c$ into the hundreds that is necessary for PMSE. Instead, they proposed that larger charged aerosols may be responsible for raising the electron Schmidt number. Figure 14 shows examples
of their calculation for positively and negatively charged aerosols. Note that an aerosol with a radius of 0.01 μm can raise $Sc$ to over 100 if ~60% or more of the total plasma charge is contained in the aerosols.

The question of whether the aerosols are charged positively or negatively and how much either way is the total plasma charge is contained in the aerosols.

0.01 μm can raise $Sc$ to over 100 if ~60% or more of charged aerosols. Note that an aerosol with a radius of 0.1 μm. Thus 0.01-1 μm particles would have to be somewhat more numerous than the electrons, while the 0.1 μm aerosol number density would still need to be a good fraction of the electron density for them to raise $Sc$ significantly. Owing to the limitation imposed by the total available water from which to create these particles, the former condition is likely, but the latter is not. In addition, a large number of 0.1 μm particles would be observable by the naked eye, and we know that visible clouds are much more rare than PMSE. Thus the raised Schmidt number theory of PMSE would seem to favor a large number of small, subvisible aerosols.

On the other hand, Havnes et al. [1990] have suggested that impurities in the ice, such as metals from meteor vapor, could substantially lower its photoelectric work function so that it may become highly positively charged. If this is true, the number density requirement for Schmidt number enhancement can be relaxed, since more charge is carried per aerosol. Observations and laboratory experiments are badly needed to shed more light on this area.

Charged aerosols have also been implicated in causing the electron density bite-outs often observed by rocket probes near the summer mesopause [Bjorn et al., 1985]. The idea is that a thin layer of small, subvisible aerosols scavenges electrons to form the sharp depletion. Reid [1990] showed that because of the limitations of total available water, ice particles with radius ~0.01 μm and number density of the order of the electron density were required to form the deep bite-outs. These numbers are consistent with the enhancement of coherent radar echoes via a raised Schmidt number with charged aerosols described above, an agreement that is further supported by the frequent observation of electron density depletions in the region of PMSE.

There is some evidence, however, for the existence of positively charged aerosols. Havnes et al. [1990] pointed out that dc Langmuir probes, which are mounted on rockets to measure the electron density fluctuations, can be fooled into observing a depletion if positively charged particles too massive to be deflected by the positive device potential ram into the probe. This is one explanation for the observation of Inhester et al. [1990] in which a Langmuir probe and an RF capacitance probe carried on the same rocket showed a large discrepancy in the electron density depletion region: the former measured a much deeper bite-out than the latter, which would not have been affected by a ram current of heavy positive particles. On the other hand, the bombardment of the Langmuir probe by massive aerosols may be energetic enough to trigger secondary electron emission, which would also explain the apparent depletion.

We have examined the effect of charged aerosols on coherent radar scatter through the turbulent scatter theory mainly because a quantitative approach has been established. Returning to the observations, however, it is apparent that Fresnel-type scatter also plays an important role in PMSE. A reduced diffusion theory analogous to the one outlined above will need to be developed, as it is intuitively clear that the retarding of electron diffusion will enhance the likelihood that sharp gradients in their density will be maintained.

**UHF PMSE**

In addition to increasing radar scatter by reducing the electron diffusivity, charged aerosols can also enhance radar scatter by introducing inhomogeneities in the ambient ionization solely due to their own charge. Historically, this phenomenon was first discussed in the context of plasma waves in dusty plasma and was called “transition scattering” [Tsytovich et al., 1989]. It was then extended to the case of electromagnetic wave scatter [Bingham et al., 1991] and was proposed as a mechanism for PMSE by Havnes et al. [1990]. Hagfors [1992] and La Hoz [1992] developed a parallel theory using the Debye-Huckel “dressed particle” approach, which was one of the bases for the development of incoherent scatter theory. For lack of an established nomenclature, we shall call this scattering mechanism “dressed aerosol scatter.” Because of its limited capacity to enhance radar scatter above the incoherent scatter level and its nondependence on radar wavelength, dressed aerosol scatter is relevant only for UHF radars where isotropic turbulence and Fresnel scatter drop off rapidly with decreasing scale size.

The idea is as follows: statistically, a charged aerosol would be surrounded by a “sphere” of surplus or deficit (corresponding to a positively or negatively charged particle) of electrons with a characteristic length scale given by the plasma Debye length, $\lambda_p$. The Debye length is a standard measure of the dis-
Figure 15. The turbulent radar reflectivity model of Driscoll and Kennedy [1985] for two values of Schmidt number $Sc$ (solid lines) [Cho, 1993]. The values of electron density, electron gas scale height, and turbulence energy dissipation rate used are listed at the top. Note the tremendous leverage that $Sc$ has in extending the radar reflectivity to shorter scales. The vertical bars indicate the range of reflectivities recorded by radars at three different frequencies. The horizontal dotted line indicates the incoherent scatter level for the given electron density. The horizontal dashed line indicates the level of dressed aerosol scatter due to a presence of charged aerosols of number density $10 \text{ cm}^{-3}$ and charge number 100.

tance up to which the plasma shields the electric potential of the charged aerosol. If $\lambda_D \ll \lambda_R$, where $\lambda_R$ is the radar wavelength, then the Debye sphere will respond in phase, therefore leading to an increase in the scattered power per aerosol over that of incoherent scatter by a factor of roughly $|Z|^2$ [Bingham et al., 1991], where $Z$ is the charge number per aerosol. Thus for an electron density that is held constant, the resultant scattering increases roughly by $|Z|$ over incoherent scatter. A more quantitative result depends crucially on the ratio of the average distance between aerosols to the plasma Debye length [de Angelis et al., 1992; La Hoz, 1992]. If the ratio is large compared with 1, then the aerosols can be considered independent, and the resulting enhancement of radar scatter can become large. However, if the ratio is smaller than 1, then the mutual interactions between the aerosols weaken the enhancement. Hagfors [1992] has shown that in the latter case, radar scatter from charged aerosols is not likely to rise above incoherent scatter levels. Thus the mechanism favors a small number of highly charged aerosols over a large number of particles with low charge. In turn, this argues for positively charged aerosols.

For the strongest PMSE event observed at 933 MHz, Cho et al. [1992a] estimated that $|Z| \sim 100$ would be necessary for the dressed aerosol scatter effect. A weaker event measured at 1.29 GHz requires $|Z| \sim 10$ [Cho et al., 1992b]. There is only a small difference between the effect of positively and negatively charged aerosols, but such high per-particle charges can probably only be achieved on the positive side through photoemission, which requires the reduction of the ice work function by metallic contamination. Such a condition may be met only intermittently, which may explain the relative rarity of PMSE at UHF.

Cho et al. [1992a] also showed that the resulting spectral width of dressed aerosol scatter from charged aerosols to decrease rapidly with the particle radius. The very narrow PMSE spectra observed by the EISCAT UHF radar [Röttger et al., 1990b] could thus be interpreted as an effect of charged aerosols.

We present Figure 15 as a summary of the effects of charged aerosols on mesospheric radar scatter. The range of radar reflectivities measured by three radars, CUPRI, EISCAT VHF, and EISCAT UHF, are displayed as vertical bars. The solid curves are the reflectivity values calculated from isotropic turbulent scatter theory. Note the extension of radar scatter to higher wavenumbers as the Schmidt number is increased from 1 to 100, i.e., when charged aerosols are present to reduce the electron diffusivity. The dotted horizontal line indicates the level of incoherent scatter for the electron density given at the top of the figure. The dashed horizontal line corresponds to the increased reflectivity due to the dressed aerosol effect. From Figure 15 it appears more likely that the UHF...
PMSE is caused by dressed aerosol scatter. However, an extreme increase in turbulent scatter due to very large, NLC-size aerosols, corresponding to Sc ~ 1000, cannot be ruled out. Although it is not depicted in this figure, we note that reduced diffusion due to the presence of aerosols will also increase the survival chances for sharp ledges in the electron density profile which can result in Fresnel scatter.

DISCUSSION AND SUMMARY

The uniquely large radar cross sections in the polar summer mesosphere occur as a result of anomalously increased structuring in the electron density at the radar Bragg scales. Experimental work using simultaneous rocket and radar measurements was the first step taken in understanding the nature of PMSE. The next step was explaining the creation and maintenance of the small electron density structures, and we have a strong candidate in the reduced diffusivity (raised Schmidt number) theory of charged aerosols. It appears that this mechanism has enough leverage to account for the observed VHF radar cross sections by itself. Small, subvisible ice particles with low negative charge and number density of the order of the electron density are the likely culprits, especially since they can also cause the electron density depletions and bite-outs often observed simultaneously with PMSE. Another factor favoring small aerosols is that the average height of PMSE (86 km) seems to be markedly above that of noctilucent clouds (83 km) [Gadsden and Schröder, 1989] which is where the largest particles settle owing to sedimentation. Consequently, with regard to the correlation or lack thereof between PMSE and NLC sightings, one would not expect such close coincidences, since the presence of NLC may indicate that most of the available water has been tied up by very large particles.

Even though the reduction of electron diffusivity can account for large radar reflectivities, other characteristics of PMSE require further inquiry. The idea that two distinct generation mechanisms (turbulent and nonturbulent) create VHF PMSE, first conjectured by Kelley and Ulwick [1988], has gained much support from the recent results of NLC-91 [Lübken et al., 1993; Cho et al., 1993; Ulwick et al., 1993]. The newly proposed generation mechanisms of electron density structures, especially gravity-wave-produced local temperature minima that may create sharp electron depletions through enhanced recombination or aerosol scavenging, should be investigated, as they may be important in explaining such features as aspect sensitivity and the lamination of thin scattering layers. The relative effects of stratification and turbulence combined with reduced electron diffusivity on the morphology of the echo layers also need more study. Interferometry and multibeam techniques can be used more extensively to probe the layer structures. Aspect sensitivity should be measured at frequencies higher than 50 MHz.

PMSE observed by the UHF radars can be explained either by reduced diffusion coherent scatter (by pushing the various parameters to the upper limits) or by dressed aerosol scatter. In either case, highly charged aerosols are necessary (in the first case, large size also), and thus one would not expect a high frequency of occurrence. A very fine spectral resolution UHF radar mode should be used to look for the extremely narrow spectra that are predicted for dressed aerosol scatter. Also, a simultaneous, collocated observation with radars of at least three frequencies (two VHF and one UHF) would be highly desirable in determining the exact frequency dependence of PMSE.

Assuming that the dependence of PMSE on charged aerosols is indeed crucial, then one of the great unknowns is the compositions, sizes, numbers, shapes, orientations, and charge states of the aerosols. Optical methods such as spectrophotometry and polarimetry have given us some information regarding the larger particles, and rocket-borne mass spectrometers have been successful in measuring ion compositions. However, we know very little about the intermediate-size regime, including meteoric dust and embryonic ice cloud particles, and the charge states of the aerosols are virtually unknown. Collection of aerosols by rockets has been attempted several times but without very conclusive results [see Gadsden and Schröder, 1989]. More clever in situ experiments need to be devised in order to further our understanding of these particles. Unfortunately, laboratory results are also very sparse, as the summer mesopause is an especially complex region to imitate.

The initial data from the recently installed VHF radar on the Palmer Peninsula in Antarctica surprisingly showed no signs of PMSE (Balsley et al., 1993). Also, satellite observations of polar mesospheric clouds, which are believed to be basically the same phenomenon as the ground-observed NLCs, indicate brighter clouds in the north than in the south [Olivero and Thomas, 1986], which implies a warmer or drier mesopause in the south. Could the asymmetry between the Arctic and Antarctic mesopause be so great as to introduce a PMSE gap? This is an intriguing question with implications concerning the global circulation, coupling between atmospheric regions, and radiative balance that require further exploration.

One final note concerning global change: NLC observations are absent from the historical record before 1885, and systematic observations in the last 2 decades have revealed an increase in their frequency of occurrence [Gadsden, 1990]. If the long-term increase in cloud formation is real, it is probably due to (1) increase in the water vapor or (2) decrease in the temperature. Thomas et al. [1989] opted for the first pos-
sibility and argued that the anthropogenic increase in methane gas (the oxidation of which in the stratosphere is an important source of water in the middle atmosphere) was responsible for the cloud increase. Gadsden [1990] noted that the second possibility was just as likely. We point out, however, that Roble and Batchelor, G. K., Small-scale variation of convected quantities like temperature in a turbulent fluid, 1, J. Fluid Mech., 5, 113, 1959.


Franke, S. J., J. Röttger, C. LaHoz, and C. H. Liu, Frequency domain interferometry of polar mesosphere sum-


Hall, C., Modification of the energy-wavenumber spectrum for heavy proton hydrates as tracers for isotropic turbulence at the summer mesopause, J. Geophys. Res., 95, 5549, 1990.


Reid, I. M., P. Czechowsky, R. Rüster, and G. Schmidt,


