JGR Space Physics

RESEARCH ARTICLE

10.1029/2018JA026171

Statistical Modeling of the Coupled *F*-Region Ionosphere-Thermosphere at High Latitude During Polar Darkness

G. D. Dorrian[1], A. G. Wood[1], A. Ronksley[1], A. Aruliah[2], and G. Shahtahmassebi[1]

[1]School of Science and Technology, Nottingham Trent University, Nottingham, UK, [2]Department of Physics and Astronomy, University College London, London, UK

Key Points:

- Linear modeling of variability in the high-latitude coupled ionosphere-thermosphere is a more accurate predictor than climatology
- Univariate linear models demonstrate the relative influences of different geophysical proxies on ionosphere-thermosphere variability

Correspondence to:
G. D. Dorrian,
gareth.dorrian@ntu.ac.uk

Citation:
Dorrian, G. D., Wood, A. G., Ronksley, A., Aruliah, A., & Shahtahmassebi, G. (2019). Statistical modeling of the coupled *F*-region ionosphere-thermosphere at high latitude during polar darkness. *Journal of Geophysical Research: Space Physics*, *124*. https://doi.org/10.1029/2018JA026171

Received 8 OCT 2018
Accepted 24 DEC 2018
Accepted article online 18 JAN 2019

Abstract Statistical models have been developed for predicting the behavior of the coupled high-latitude ionosphere-thermosphere system. The modeled parameters were the *F*-layer peak electron density, plasma structuring, ion temperature, neutral temperature, and the difference between these temperatures, which is a key term in the Joule heating equation. Ionospheric measurements from the European Incoherent Scatter Svalbard Radar and neutral atmosphere measurements from the colocated University College London Fabry-Perot Interferometers have been made across a solar cycle. These data were all acquired during nighttime conditions as the observations with the Fabry-Perot Interferometers are restricted to such times. Various geophysical proxies were tested to represent the processes that influence the modeled parameters. The dominant geophysical proxy for each modeled parameter was then determined. Multivariate models were also developed showing the combinations of parameters that best explained the observed variability. A comparison with climatology showed that the models give an improvement in every case with skill scores based on the mean square error of up to 0.88.

Plain Language Summary The upper atmosphere of the Earth is a mixture of partially ionized plasma (the ionosphere) and neutral gases (the thermosphere). The plasma is comprised of charged particles which are subject to electromagnetic forces, whereas the neutral thermosphere is not. These two populations are the coupled ionosphere-thermosphere. Plasma density in the ionosphere is controlled by plasma production mechanisms such as solar ultraviolet illumination, plasma loss mechanisms such as collision induced recombination, and transport mechanisms such as variability in the geomagnetic field. Of particular interest is the process of heat transfer from ionospheric plasma to the neutral thermosphere, so-called *Joule heating*. Combined with geophysical information, such as the *Kp* index, variability in the interplanetary magnetic field, and season, this study presents a series of statistical linear models that rank various physical effects in terms of how strongly they influence the behavior of the coupled ionosphere-thermosphere at high latitude during polar darkness. These models also show which combinations of physical effects enable predictions to be made of observed ionosphere-thermosphere behavior, to a greater accuracy than by a climatological approach. This study was made possible using ionosphere data collected from the European Incoherent Scatter Svalbard radar with neutral thermosphere observations from a colocated Fabry-Perot Interferometer.

1. Introduction

Global Navigation Satellite System (GNSS) has become a ubiquitous technology, with 5.8 billion devices such as mobile phones in use in 2017 worldwide, and by 2020, this is forecasted to increase to 8 billion—more than one device per person on the planet. However, it is subject to various vulnerabilities, one of which is ionospheric scintillation (European GNSS Agency market report of 2017). The dependence of our society on such technological systems means that a deeper understanding of the underlying physical processes is urgently needed (Hapgood, 2017).

The ionosphere is a weakly ionized plasma in the Earth's atmosphere extending from an altitude of ~60 to ~1,000 km, where it merges with the Earth's outer environment. In the polar cap this plasma is structured on a wide range of horizontal spatial scale sizes (Tsunoda, 1988). Large scale structures with a horizontal extent of tens to hundreds of kilometers exhibit variation with season, solar cycle, geomagnetic activity, solar wind conditions, time of day, and location (Hargreaves, 1992). These large-scale structures can also cause

small-scale irregularities, which grow due to the irregularity wave cascade process driven by the gradient drift instability and/or turbulent processes (Burston et al., 2010).

These smaller-scale structures can cause rapid phase and amplitude fluctuations in the GNSS signals; a phenomenon known as scintillation (Kintner et al., 2007). A direct connection between gradients in the total electron content (TEC) at the edge of a plasma stream and scintillation has been observed (Mitchell et al., 2005) and plasma structuring caused by auroral precipitation has been linked to the loss of signal lock by a GNSS receiver (Elmas et al., 2011). Statistical studies have shown an agreement between scintillation and the asymmetric distribution of polar cap patches around magnetic midnight (Spogli et al., 2009) and that auroral emissions correlate with GNSS signal scintillation (Kinrade et al., 2013).

The present study focused upon the *F* region, which extends in altitude from approximately 150–600 km and is itself broken down into two distinct regions, known as the *F*1 and *F*2 regions. The *F*2 region contains the highest density of free electrons in the ionosphere, which peaks at approximately 10^{12} electrons per cubic meter. Neutral particle densities in the thermosphere, at the altitude of the *F*-region peak, are approximately 10^{14} m^{-3}. At high latitudes, there is an approximately linear correlation between density of atomic oxygen, a major constituent of the neutral atmosphere at this altitude, and *F*10.7 solar flux illumination (Vickers et al., 2014). Variations in thermosphere are strongly dependent on both solar illumination and activity. The temperature of the ions at this altitude is variable and is dependent on many factors including latitude, time of day, and solar activity but is generally greater than the neutral thermosphere of the order of 1000 K. This difference in temperature results in the thermal energy of the plasma being transferred to the neutral atmosphere through ion-neutral collisions. The movement of ions through the resistive medium of the thermosphere therefore results in Joule heating. The Joule heating rate (Q_C) is shown in equation (1) (Banks & Kockarts, 1973; Brekke & Kamide, 1996; Cole, 1962; Schunk, 1975; Vasyliunas & Song, 2005):

$$Q_C = \sum_{\mathrm{n}} n_{\mathrm{n}} m_{\mathrm{n}} \sum_{\mathrm{i}} \frac{\nu_{\mathrm{in}}}{m_{\mathrm{n}} + m_{\mathrm{i}}} \left[3k(T_{\mathrm{i}} - T_{\mathrm{n}}) + m_{\mathrm{i}} (\overline{u}_{\mathrm{n}} - \overline{u}_{\mathrm{i}})^2 \right], \tag{1}$$

where n_{n}, m_{n}, n_{i}, and m_{i} are the number densities and masses for the neutrals and ions, respectively. The mean wind velocities of the neutrals and ions are given as $\overline{u}_{\mathrm{n}}$ and $\overline{u}_{\mathrm{i}}$, and ν_{in} is the ion-neutral collision frequency. T_{i} and T_{n} are the temperatures of the ions and neutrals, and k is Boltzmann's constant.

Ionization on the dayside is dominated by direct illumination by solar ultraviolet and X-ray radiation. The rate of ion-electron pair production is dependent on the intensity of incident ionizing radiation upon the atmosphere, the number density of particles capable of being ionized, the absorption cross section of these particles, and an ionization efficiency term. This combination of terms is the Chapman production function (Chapman, 1931). The maximum rate of production reaches a maxima where the optical depth of the atmosphere approaches unity, typically in the *F* region (Hargreaves, 1992).

Ion-electron pair production on the nightside is dominated by ionization of the neutral atmosphere by high energy particles propagating along the Earth's magnetic field lines and entering the atmosphere from above. The interplanetary magnetic field (IMF) flux is transferred to the magnetotail via magnetic reconnection and in regions above the geomagnetic poles, to which the IMF flux can be directly coupled into the Earth environment (e.g., Dungey, 1961; Gonzalez et al., 1994). Particle precipitation along these field lines is time variant and is particularly intense during geomagnetic storms and substorms (Brekke, 1997). Reconnection occurs on a continual basis, driven by persistent solar wind interaction with the dayside magnetosphere. This also results in particle precipitation directly from the solar wind into the dayside ionosphere (Frey et al., 2003). Plasma production is further stimulated at all latitudes by the ubiquitous incidence of cosmic rays upon the Earth's atmosphere (Velinov, 1968).

Plasma produced by photoionization can be drawn in to the polar cap. Foster (1984) first characterized this process using average maps of both ionospheric convection flows and electron density derived using data from the Chatinka incoherent scatter radar. One type of large-scale electron density enhancement in the *F* region is known as a polar cap patch. These were defined by Crowley (1996) to have a horizontal extent of at least 100 km and plasma densities at least twice that of the surrounding background ionosphere hence with a patch-to-background ratio of $\geq$2. Subsequent observations indicated that patches could be transported over large distances. An individual patch was followed for more than 3,000 km from the center of

the polar cap to the poleward edge of the nightside auroral oval (Weber et al., 1986). The Utah State University Time-Dependent Ionospheric Model has been used to show that the plasma can traverse the entire polar cap (Sojka et al., 1994). Patches have been observed drifting out of the polar cap (Pedersen et al., 2000) and being reconfigured to form another type of plasma density enhancement known as a boundary blob (Pryse et al., 2006).

Plasma loss is due to dissociative recombination, the rate of which is governed by the reactions of O^+ with N_2 and O_2. These reaction rates depend upon the recombination coefficients, the plasma density, and the density of the neutral species, and many of these also depend upon the temperature of the reactants, as shown in reaction rates given by McFarland et al. (1973), St.-Maurice and Torr (1978), and Hierl et al. (1997).

Polar cap plasma exhibits seasonal variation. This was first observed using the Chatinka incoherent scatter radar, where dayside plasma densities drawn into the polar cap were 3 times greater in winter than in summer (Foster, 1984). A similar seasonal pattern of higher winter densities had previously been reported at midlatitudes by Rishbeth and Setty (1961) and Wright (1963). The authors attributed this to the seasonal anomaly where higher temperatures caused upwelling of the thermosphere in the summer hemisphere. This led to lower $O/(N_2)$ and $O/(O_2)$ atomic/molecular ratios, which increased chemical recombination and consequently decreased the plasma density.

Wood and Pryse (2010) observed the seasonal change in the patch-to-background ratio, a measure of the variability of polar cap plasma, above northern Scandinavia. Patch-to-background ratios were up to 9.4 ± 2.9 in winter and up to 1.9 ± 0.2 in summer. Modeling suggested that this difference was primarily due to variation in the chemical composition of the atmosphere, which, in summer, both reduced the electron densities of the plasma drawn into the polar cap on the dayside and enhanced plasma loss by recombination. A secondary factor was the maintenance of the background polar ionosphere by photoionization in summer.

Polar cap plasma is also influenced by the IMF conditions. A statistical study of patches close to the geomagnetic pole showed that these structures were primarily associated with times when IMF B_Z was negative (McEwen & Harris, 1996). This was attributed to the convection pattern under these conditions being favorable for the antisunward cross-polar transport of plasma.

The solar cycle is another influence on the patch-to-background ratio. Simulations predicted that patch lifetimes would be longer at solar maximum than at solar minimum, with this difference attributed to changes in the solar extreme ultraviolet (EUV) radiation intensity (Sojka & Schunk, 1987). Around solar maximum winter nightside patches were observed using the Sondrestrom incoherent scatter radar (Pedersen et al., 1998). The majority of patch-to-background ratios were between 2 and 4, although values as high as 9 were recorded.

Production of ionization by particle precipitation occurs at high latitudes. This occurs where ions or electrons collide with the neutral atmosphere. This mechanism is not seasonally dependent and, in the dark winter ionosphere, the absence of sunlight means that particle precipitation is the dominant production mechanism. A modeling study by Millward et al. (1999) showed that the altitude at which the plasma production rate is greatest varies with the energy of the precipitating particles. This showed that higher energy electrons penetrate to lower altitudes; however, it is the lower energy electrons that are primarily responsible for energy deposition at greater altitudes. Precipitating particles deposit energy in the atmosphere. This can increase plasma and neutral temperatures, increasing the scale height of the atmosphere and driving changes in the neutral composition which can, in turn, alter plasma production and recombination rates.

Burns et al. (2004) suggested that variation in the neutral composition could also contribute to the formation of patches. A modeling study showed that, in summer, neutral species were drawn antisunward with the plasma due to collisions. Collisions with neutrals already in the polar cap resulted in Joule heating. This caused a pressure gradient, which transported molecular rich air from polar cap to daytime midlatitudes, the source region of this plasma. This resulted in enhanced recombination and broke up the tongue of ionization (TOI).

Airglow is a faint and diffuse emission of light, which can be detected from the ground, which arises from the deexcitation of species in the neutral atmosphere. At the altitude of the *F* region, a faint red light airglow emission from deexcitation of atomic oxygen at 630 nm can be visible during the hours of darkness (Hays et al., 1978).

The dependence of ionospheric plasma structures on the thermosphere has never been fully established, despite the well-understood ionosphere-thermosphere coupling mechanism. The thermosphere is also a key parameter in the Joule heating equation, due to thermal energy transfer from the ionosphere to the thermosphere by way of collisions between ions and neutrals. In this paper, we present a statistical comparison of plasma and neutral atmosphere observations made using independent colocated ground-based instruments at high latitudes over more than a solar cycle. We apply linear modeling to determine the relative importance of the various physical processes that cause variability in the ionosphere and neutral atmosphere (section 5), and create predictive models (section 6). We compare the goodness of fit of these models with a climatological analysis (section 6) to demonstrate the efficacy of this approach.

2. Instrumentation

The presence of free electrons in the ionosphere results in the scattering of radio waves. Instruments such as the European Incoherent Scatter (EISCAT; Rishbeth & Williams, 1985) radars have long been used a means to probe ionospheric characteristics. In this study, we use measurements of electron density, electron temperature (T_e), and ion temperature with the 32- and 42-m antennas of the EISCAT Svalbard Radar (ESR; Wannberg et al., 1997). Both antennas are colocated at 78.2°N, 16.0°E, 75.2° MLAT, and 112.9° MLON, at Longyearbyen on Svalbard.

The ESR is approximately colocated with a Fabry-Perot Interferometer (referred to as the FPI) at the Kjell Henriksen Observatory, operated by University College London (e.g., Aruliah et al., 2004), which can measure parameters of the neutral thermosphere above the observatory site by measuring emission intensity from atomic oxygen (I). The altitude for peak red line emission intensity occurs at ~240 km, which makes the Svalbard FPI particularly suitable for probing the behavior of the neutral atmosphere in the same altitude range as the *F*-region ionosphere. The thermosphere parameters of temperature and wind velocity can be resolved by the FPI to within uncertainties of ±90 K, and ±10 m/s, respectively. The FPI operates over a bandwidth of 1 nm centered at 630.2 nm; airglow at these wavelengths is approximately 100 times fainter than the brightness of the full moon (Hargreaves, 1992) and therefore requires the sky to be dark in order to obtain useable data.

The IMF carried by the solar wind was monitored at the Lagrangian L1 point by the Advanced Composition Explorer spacecraft (Zwickl et al., 1998). Geophysical data also used included *F*10.7 solar flux, the *Kp* index, the *Dst* index, and a geometric function for season (day of year).

3. Modeling Method

Linear models are used in this analysis. This approach can be applied in situations where there are numerous, competing physical processes. A number of explanatory variables act as proxies for the geophysical processes. This method can determine the relative importance of these explanatory variables and which combination of them best predicts a physical parameter, such as the electron density at the *F*-layer peak.

3.1. Dependent Variables

In this study, five dependent variables were modeled in turn. These were the electron density at the *F*-layer peak (F_{Ne}), the structuring ratio (F_{str}), the ion temperature at an altitude of 240 km (T_i), the neutral temperature at the same altitude (T_n), and the difference in these temperatures ($T_i - T_n$). F_{Ne} is the electron density at the *F*-layer peak, and F_{str} is a proxy for the amount of plasma structuring at the same altitude. F_{str} is the ratio of the mean value of the largest 25% and the smallest 25% of the electron density measurements in 2-hr interval. This is a modification of a structuring ratio applied by Wood and Pryse (2010). T_i was observed at an altitude of 240 km as this coincides with the altitude of peak red line emission detected by the FPI, from which T_n is inferred. This allowed us to also model $T_i - T_n$, an important term in the Joule heating rate equation (equation (1)). The selection of dependent variables was chosen to investigate the interaction of the ionosphere and the neutral thermosphere, with particular regard to Joule heating. It is important to note that T_i includes information from both background ionospheric plasma and plasma structures. By contrast the dependent variables F_{Ne} and F_{str} are, by their nature, highly dependent on the presence of plasma structures.

All of these variables were given in 2-hr bins advancing in 2-hr increments, with each bin starting at 00 UT, 02 UT, 04 UT, and so forth. Given the illumination constraints specified above, the total data set extends, discontinuously, from 2001 to 2015 and hence provides partial coverage of both Solar Cycles 23 and 24. There are approximately 1,600 data points across the entire set, totaling some 3,200 hr of observations under varying space weather and geophysical conditions. To ensure a sufficiently large number of data points to produce statistically significant results these data were further binned into 6-hr windows centered on 12 magnetic local time (MLT; 06–12 UT), 18 MLT (12–18 UT), 00 MLT (18–00 UT), and 06 MLT (00–06 UT). These time sectors are referred to as noon, dusk, midnight, and dawn, respectively.

A series of histograms are presented in Figure 1; one for each dependent variable data set across all time sectors. Each histogram is fitted with two colored curves indicating the best fit to the dependent variable data for two different distribution types, normal (black curve), and lognormal (blue curve).

A greater range of distributions (not shown) were trialed. These were gamma, log-gamma, exponential, and Weibull. No physical hypothesis was implied in the trialing of these commonly used distributions; the aim was simply to establish which of them was most appropriate in each case. The X^2 error for the different distribution types were calculated. Log-normal was found to be the most appropriate distribution type for all dependent variables, except for T_i, which was normal.

3.2. Single-Term and Multiterm Models

Univariate statistical models can determine the statistical significance of the relationship between the dependent variable and the explanatory variable used within that model. If all explanatory variables are tested in turn, then these can be ranked in order of significance and the relative importance of the driving processes that these variables represent can be determined.

Such an analysis was undertaken for the five dependent variables (F_{Ne}, F_{str}, T_i, T_n, and $T_i - T_n$) and the results are shown in Tables 3–7, in section 5. In each case the single-term analysis was performed for the whole day model and then for each of the four time sectors individually, giving a total of five trials per dependent variable-explanatory variable pair.

Whereas the single-term statistical models determine the relative importance of each explanatory variable, a multiterm (multivariate) model yields the *combination* of explanatory variables, which together account for the behavior of the dependent variable. This is a nonhypothesis driven, exploratory method, whereby for each dependent variable, the initial model is comprised of all available explanatory variables, as terms in the equation (with the exception of the dependent variable itself).

The multiterm method assumes that each explanatory variable used in the model is linearly independent of the others. For each pair combination of explanatory variables in the model, a cross-correlation check was performed. If any two were found to have a correlation of $\geq|0.2|$, then the explanatory variable in the pair with the least statistical significance (based on the highest p value) was removed from the equation. For example, during the multiterm method a statistically significant level ($p \leq 0.05$) of correlation existed between $F10.7$ solar flux, and the red line emission intensity detected by the FPI. If the p value of the $F10.7$ was, say, 0.02, and the p value of I was 0.001, then the $F10.7$ term was removed from the equation, as smaller p values indicate increasing statistical significance. The model was then recomputed with only the remaining explanatory variables. There were a small number of pairs of explanatory variables with statistically significant correlations of $\leq|0.2|$. In general these arose due to sampling issues and so these were not excluded, for example, in the case of F_{Ne} and IMF $B_{Z(SD)}$ in the dawn sector, correlation was found to be 0.14 (p value of 0.041). After removal of outlier points, correlation dropped to 0.0158, and the p value increased to 0.82.

Computing the multiterm model for each dependent variable was an iterative process. Once no linearly dependent pairs of explanatory variables remained in the initial list of terms, the model was computed. Then the least statistically significant explanatory variable in the equation was removed, and the model was recomputed. This process was repeated until only terms which a statistical significance (p value) of $\leq$0.05 remained.

The general form of the model, for a dependent variable with a lognormal distribution is

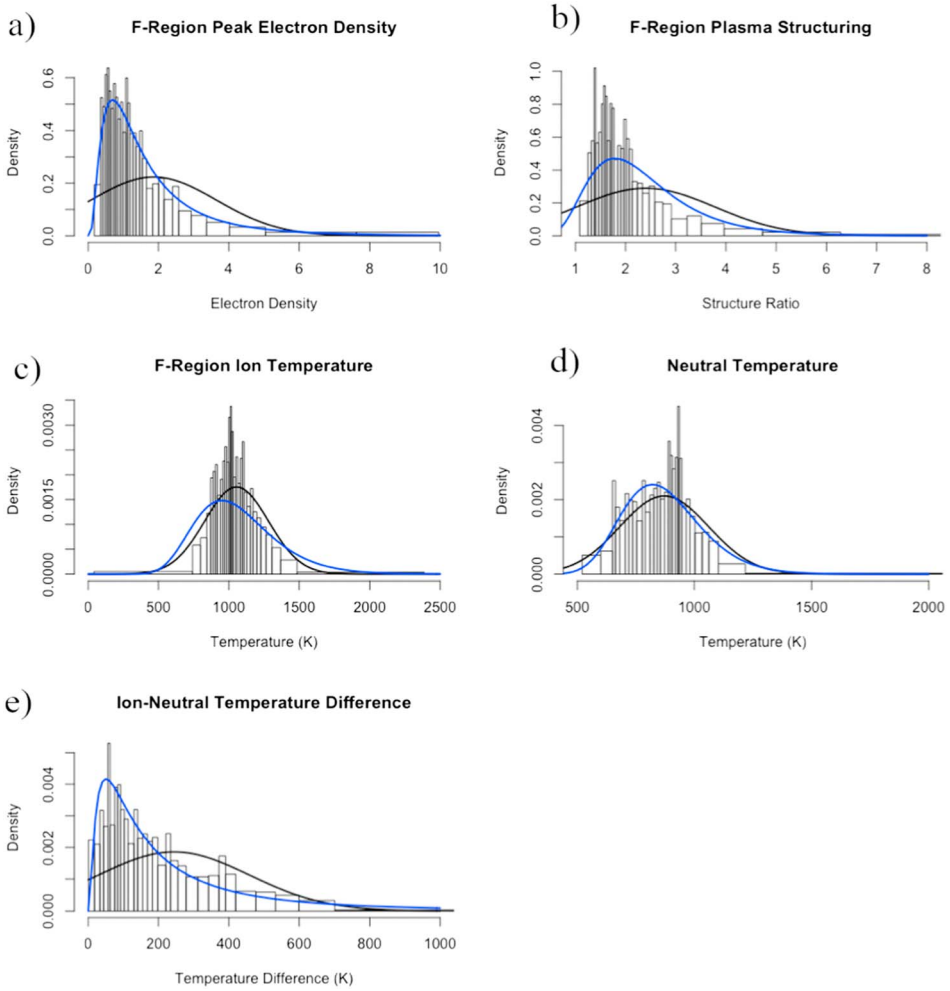

Figure 1. Panels a–e show histograms and fitted distribution curves for the data for F_{Ne}, F_{str}, T_{i}, T_{n}, and $T_{\mathrm{i}} - T_{\mathrm{n}}$, respectively. Black is normal, and blue is lognormal. For all dependent variables except T_{i}, it was found that a lognormal distribution was the best fit to the data.

$$\ln(y) \sim N(\mu, \ \sigma^2), \tag{2}$$

where y is the response variable and μ and σ^2 are the mean and variance of the Gaussian distribution. Therefore, the linear model can be defined as follows

$$E(\ln y) = \beta_0 + \beta_1 \ x_1 + \ldots + \beta_p x_p, \tag{3}$$

where $E(\ln y) = \mu$ and $\beta_0, \ldots, \beta_p$ are parameters of the model and $x_1, \ldots, x_p$ are explanatory variables.

For a case specific example, consider that after the initial removal of linearly dependent pairs, it is found that the model equation contains the variables $F10.7$, U_{merid}, I, and T_{i}. If the p values associated with these variables are 0.25, 0.15, 0.01, and 0.005, respectively, then $F10.7$ would be removed given that it has the highest p value (0.25) and the model recomputed. If the p values associated with the remaining terms U_{merid}, I, and T_{i}, are 0.07, 0.1, and 0.001, respectively, then the process would repeat again, this time removing the I term (red line emission intensity). If the remaining terms U_{merid} and T_{i} now have p values of 0.01 and 0.02, respectively, then the process would be complete, and the final equation of the model would be of the form

$$E(\ln y_l) = \beta_0 + \beta_1 U_{\mathrm{merid}_l} + \beta_2 T_{i_l} \tag{4}$$

for a given dependent variable at the lth index in the data set. β_1 and β_2 are the coefficients for that explanatory variable (at index l), and β_0 is the intercept.

Table 1
List of Variables Used in the Modeling Analysis

Variable abbreviation	Variable description	Data source
T_e	Two-hour averaged electron temperature, 240-km altitude (K)	ESR
T_i	Two-hour averaged ion temperature, 240-km altitude (K)	ESR
T_n	Two-hour averaged neutral temperature, 240-km altitude (K)	FPI
F_{Ne}, E_{Ne}	Two-hour averaged electron density in the *F* region or *E* region, respectively	ESR
F_{str}, E_{str}	Structuring ratio calculated across 2 hr in the *F* region or *E* region, respectively	ESR
hmF2, hmE	Height of maximum electron density, *F*2 region and *E* region, respectively	ESR
B_x, B_y, B_z, $B_{x(SD)}$, $B_{y(SD)}$, $B_{z(SD)}$	IMF *X*, *Y*, and *Z* components (nT) in GSM coordinates (averages and standard deviations)[a]	ACE
C, C_{SD}	IMF clock angle (°), (average and standard deviation)[a]	ACE (calculated)
Kp	*Kp* index	*Kp* index
I	Red line emission intensity (kR; 240 km)	FPI
U_{zon}, U_{zen}, U_{merid}	Two-hour averaged wind velocity, zonal, zenithal, and meridional	FPI
$\lvert U_{zon}\rvert$, $\lvert U_{zen}\rvert$, $\lvert U_{merid}\rvert$	Two-hour averaged absolute wind velocity, zonal, zenithal, and meridional components	FPI
*F*10.7	*F*10.7 solar flux	Standard *F*10.7 solar flux
$T_i - T_n$	Difference between T_i and T_n	ESR and FPI
$X(Y)$	Explanatory variable *X*, lagged by *Y* hours	Various
$X(-Y)$	Explanatory variable *X*, lagged by $-Y$ hours	Various
DOY	Day of year (geometric function)	Calculated
T_{geo}	Geometric time of day function	Calculated

Note. Some of these variables, such as T_i, are used as both explanatory and dependent variables, although not for the same model. ACE = Advanced Composition Explorer; ESR = European Incoherent Scatter Svalbard Radar; FPI = Fabry-Perot Interferometer; GSM = geocentric solar magnetospheric.
[a]The time taken for information to propagate from the location of the ACE spacecraft (L1) to the ionosphere is typically more than 1 hr but less than 2 hr. Therefore, only data from the ACE spacecraft observed from 2 hr before the first ESR observation until 1 hr before the final ESR observation were considered in each case.

Physically speaking, the end result of this process is a series of equations consisting only of physical influences that, in combination, explain (to some degree) the behavior of the dependent variable. It is for this reason that we call this method "exploratory" and "non-hypothesis driven"; we make no hypothetical predictions about the final outcome and just observe which explanatory variable combinations emerge from the process. This approach of removing variables sequentially is commonly performed in the technique of generalized linear modeling (McCullagh & Nelder, 1983). A special case of generalized linear modeling is linear modeling, and this same iterative approach is used here. The reason why we do not generalized linear modeling in this study is because, as shown in section 3.1, only nonexponential families of distributions were identified as being appropriate for the parameters under study.

3.3. Explanatory Variables

A range of explanatory variables were trialed, as shown in Table 1. A subset of these, as discussed in section 3.1, were used as dependent variables. These were also trialed as explanatory variables when modeling other dependent variables, for example, T_n was trialed as an explanatory variable for F_{str}.

The full data set was split randomly into two equal size subsets. One set, hereafter referred to as the training data, was used to produce all the predictive models and climatology. The second set, hereafter referred to as the test data, was used as an independent data set upon which the predictive models were tested.

As a further check against the unforeseen inclusion of bias in our data subsets due to specific data indices being chosen, the means of all variables used in both the training data and the test data were also calculated. In all cases, any differences were not statistically significant. It can therefore be assumed that our training and test subset data are representative of the whole set.

4. Climatology

In this section, the results of a climatological analysis are presented, which show the changes in behavior of the dependent variables T_i, T_n, $T_i - T_n$, F_{Ne}, and F_{str} throughout the day. The data for each dependent variable is averaged into the four time sectors centered on midnight, dawn, dusk, and noon MLT. Each time sector covers 6 hr of data. All data used to produce the climatology results are sourced only from the training

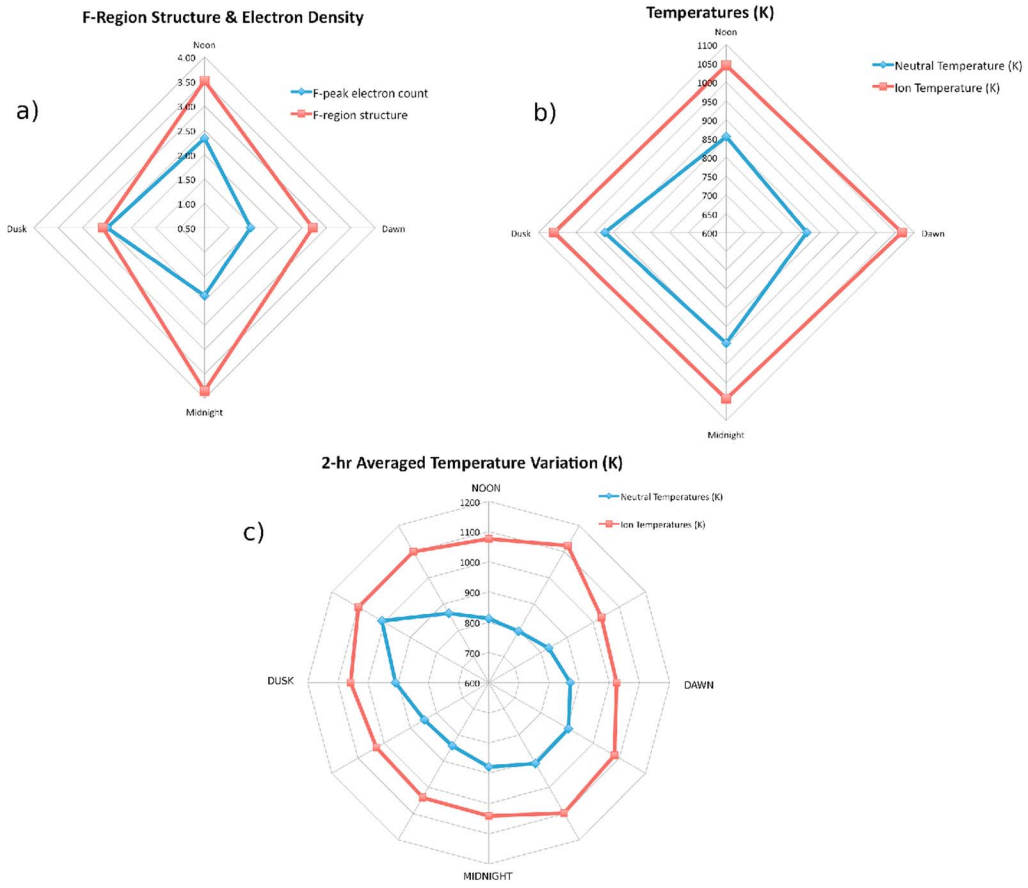

Figure 2. Climatology analysis of the dependent variables, distributed by time sector. Plot (a) shows F_{Ne} and F_{STR}, by 6-hr time sector centered at noon (top), dusk (right), midnight (bottom), and dusk (left). Plot (b) shows T_i and T_n, similarly distributed by time sector. The data volume and quality was such that the temperature data could be expressed in 2-hr time bins, centered on the times shown in plot (c). All times are magnetic local time. Neutral temperatures can be seen to reach a maxima around noon and remain elevated in the latter half of the day, despite all these data being collected during northern polar winter, which may indicate that heated neutral air is being transported over the terminator from the sunlit side of Earth. The numeric values used in the top two plots are shown in Table 2.

data. This was done to maintain consistency with the multiterm modeling procedure (section 6), against which the climatology results are compared for predictability.

Variability in T_i, T_n, F_{Ne}, and F_{str}, by time sector, is shown in the radar plots in Figure 2. Table 2 shows the data volumes, mean averages, and standard deviations (*SD*s) for the dependent variables in each time sector. It is these average values that form the baseline for the comparison with climatology (section 6).

Figure 2b shows that T_i remains fairly constant throughout the day. Table 2 shows that these values are fairly well constrained, with an *SD* of approximately 20% of the average value in each time sector. The climatology plot for T_n shows elevated temperatures in the noon and dusk sector, although the variation is smaller than the associated *SD*.

For the T_i and T_n, there were sufficient numbers of data points available in the training set to enable a higher time resolution analysis. Figure 2c shows the temperature variations broken down into each 2-hr time bin rather than the 6-hr time sector described above. More than 100 data points in each case were used to calculate the values in this plot. T_n minimizes at approximately 0800 UT, just after dawn, reaches a maximum at approximately 1200 UT, and remains elevated throughout the afternoon, with respect to the morning hours. While this may be expected under normal illumination conditions, all these data were acquired during hours of polar darkness. Ion temperature, as in the lower time resolution plots, is observed to remain fairly constant throughout the day and at all times greater than T_n.

Table 2
Numerical Values for T_i and T_n, F_{Ne}, F_{str}, and $T_i - T_n$ Acquired From Climatology, by Time Sector

	T_i	T_i *SD*	T_n	T_n *SD*	F_{Ne}	F_{Ne} *SD*	F_{str}	F_{str} *SD*	$T_i - T_n$	Data points
Midnight	1043	229	895	193	1.89	2.44	3.85	6.63	148	384
Dawn	1067	278	814	138	1.44	1.27	2.72	3.23	253	420
Noon	1044	187	855	140	2.34	2.38	3.51	11.03	189	405
Dusk	1059	252	923	192	2.50	2.53	2.59	3.78	136	397

The variability in F_{Ne} and F_{str} can also be seen in Figure 2a. Peak electron count maximizes in the noon and dusk sectors and minimizes at dawn. This behavior somewhat mirrors the behavior of T_n (Figure 2b). *F*-region structuring clearly shows maxima at noon and midnight. The noon maximum may be related to particle precipitation resulting from reconnection events at the dayside magnetopause and midnight maxima may be related to reconnection events in the magnetotail for the nightside (e.g., Frey et al., 2003; Hubert et al., 2006). In the cases of both *F* region electron structuring and concentrations, there are relatively large *SD*s (Table 2), which suggests that climatology can be improved upon as a predictive tool in this context.

5. Single-Term Model Results

Single-term models were used to determine the relative importance of the driving processes, using the method outlined in section 3.1.

In order to provide a ranking for statistical significance a score was set for each single-term model as indicated by the *p* value. A score of "1" indicates a significance at the 5% level, "2" indicates a significance at the 1% level, "3" indicates a significance at the 0.1% level, "4" indicates a significance at the 0.01% level, and "5" indicates a significance at the 0.001% level. The mean value of this score across the trials for the four time sectors for each dependent variable-explanatory variable pair was used to produce the rankings shown in Tables 3–7. The color shadings in Tables 3–7 indicate the ranking score for a given explanatory variable; pink indicates a score of 5, orange for 4, yellow for 3, green for 2, and blue for 1.

Tables 3–7 show the single-term model results for all explanatory variables where the average significance ranking was 2.0 or better. The parameter estimates for each of the five trials (the whole day model and each of the four time sectors) are shown. Parameter estimates for the time sector models are shaded yellow if they are significantly different from the associated whole day model. This indicates that the physical relationship between the dependent-explanatory variable pair shows significant variation. Parameter estimates are absent if they are not statistically significant at the 5% level, which indicates that there is no statistically significant relationship between these parameters in our data set for a given time sector.

5.1. Electron Density at the *F*-Layer Peak (F_{Ne})

The results for the electron density at the *F*-layer peak (F_{Ne}) are shown in Table 3. The most significant explanatory variable is the *F*10.7 cm solar flux, indicating that the well-known importance of solar activity dominates. It is emphasized that all these data were gathered during hours of darkness, and this suggests that transport and/or corotation brings plasma produced by photoionization from the dayside. Based on the parameter estimates, this effect is significantly smaller at dusk and larger at midnight than for the day as whole and maximizes at midnight. The intensity of the airglow emission also maximizes at midnight. As airglow can be related to polar cap patches (i.e., Lorentzen et al., 2004), which are transported from magnetic noon to midnight, this also suggests that plasma transportation has a significant effect.

Season is a highly significant explanatory variable in the noon, dusk, and midnight sectors (Table 3, green shaded). The parameter estimate for the all-day case, shows that a higher value of N_e was expected further away from midwinter. The sector by sector parameter estimates show that N_e maximized at noon, where the solar terminator is closest to the observations. N_e is lower at midnight, where plasma would need to be transported further over the polar cap and so would have more time to decay toward background values.

Table 3
The Relative Importance of the Geophysical Proxies Influencing the Electron Density at the F-Layer Peak (F_{Ne})

	F_{Ne}											
	Significance						Parameter estimates					
Dependent Variable	Overall	Dawn	Noon	Dusk	Midnight	Average	All	Error	Dawn	Noon	Dusk	Midnight
*F*10.7 cm solar flux	5		5	5	5	5	9.93E-03	4.16E-04	9.14E-03	8.98E-03	6.37E-03	1.49E-02
HmE	5	5	5	5	5	5	1.76E-05	9.56E-07	1.62E-05	1.44E-05	2.68E-05	2.20E-05
Intensity	5	5	5	5	5	5	1.45E-03	1.22E-04	1.86E-03	1.06E-03	2.45E-03	3.04E-03
E region STR v1	5	5	1	5	5	4	2.66E-01	3.31E-02	3.63E-01	1.03E-01	5.58E-01	6.16E-01
HmF2	5	5	5	1	3	3.5	2.96E-06	3.42E-07	2.96E-06	6.46E-06	1.53E-06	2.57E-06
Season	5	0	5	5	1	2.75	4.56E-01	6.66E-02		9.27E-01	8.80E-01	3.00E-01
IMF B_z (stdev)	5	2	2	2	2	2	8.60E-02	1.63E-02	8.44E-02	8.32E-02	9.73E-02	1.08E-01
Dst (average)	3	4	0	2	2	2	7.89E-03	2.16E-03	1.38E-02		1.10E-02	1.21E-02
T_i	3	0	3	3	0	1.5	3.12E-04	8.97E-05		5.31E-04	6.73E-04	
T_e	3	0	1	0	4	1.25	2.21E-05	6.63E-06		4.05E-05		1.63E-04
Zonal wind (magnitude)	5	0	0	1	4	1.25	1.35E-06	2.31E-07			1.77E-06	1.09E-06
Meridional wind (magnitude)	3	0	0	0	5	1.25	1.87E-06	4.85E-07				5.72E-06
UT	5	2	1	0	2	1.25	-2.65E-01	3.01E-02	−4.02E-01	−3.01E-01		−4.64E-01
IMF B_y (stdev)	4	0	2	2	0	1	6.16E-02	1.55E-02		7.00E-02	9.24E-02	

Note. In order to provide a ranking a score was set in each case for the significance, as indicated by the *p* value, with "1" indicating a significance at the 5% level (blue), "2" indicating a significance at the 1% level (light green), "3" indicating a significance at the 0.1% level (yellow), "4" indicating a significance at the 0.01% level (orange), and "5" indicating a significance at the 0.001% level (pink). The mean value of this score across the trials for the four time sectors for each dependent variable-explanatory variable pair was used to produce the rankings shown in the column headed "average." In the interests of concision any parameter with an average ranking score of <1 is not shown in the table. Parameter estimates for the time sector models are shaded yellow if they are significantly different from the associated whole day model. Parameter estimates are absent from Tables 3–7 if they are not statistically significant at the 5% level. IMF = interplanetary magnetic field.

Table 4
Plasma Structuring at the F-Layer Peak (F_{str})

	F_{str}											
	Significance						Parameter estimates					
Dependent Variable	Overall	Dawn	Noon	Dusk	Midnight	Average	All	Error	Dawn	Noon	Dusk	Midnight
Season	5	5	5	5	5	5	−4.74E-01	4.37E-02	−3.54E-01	−4.04E-01	−6.02E-01	−7.13E-01
E peak Ne	5	5	5	5	4	4.75	1.25E-01	7.52E-03	1.01E-01	2.28E-01	1.45E-01	7.86E-02
T_e	5	5	3	4	3	3.75	3.11E-05	4.36E-06	3.04E-05	3.68E-05	2.88E-05	9.75E-05
PCI (average)	5	4	5	2	3	3.5	1.28E-01	1.91E-02	1.66E-01	1.86E-01	9.68E-02	1.37E-01
HmF2	4	3	4	5	1	3.25	1.03E-06	2.33E-07	−1.40E-06	1.81E-06	2.64E-06	1.41E-06
IMF B_z (average)	5	3	5	1	2	2.75	−3.89E-02	5.91E-03	−3.65E-02	−4.31E-02	−2.80E-02	−4.90E-02
Kp	5	4	5	0	0	2.25	6.50E-03	1.15E-03	8.63E-03	8.74E-03		
T_i	0	0	4	0	4	2	−1.55E-05	5.99E-05		3.50E-04		−7.27E-04
Clock angle	5	1	5	0	2	2	−1.55E-01	2.95E-02	−1.11E-01	−2.21E-01		−2.26E-01
PCI (stdev)	5	2	4	0	1	1.75	3.58E-01	5.99E-02	3.23E-01	4.34E-01		3.04E-01
Dst (average)	3	0	5	0	0	1.25	−4.73E-03	1.41E-03		−1.39E-02		
Intensity	0	0	3	0	2	1.25	−1.79E-04	9.98E-05		3.68E-04		−1.00E-03
HmE	0	1	3	0	0	1	−4.81E-07	7.06E-07	−2.85E-06	3.54E-06		

Note. Ranking and color conventions as per Table 3. IMF = interplanetary magnetic field; PCI = polar cap index.

Table 5
Ion Temperatures (T_i)

	T_i											
Dependent variable	Significance						Parameter estimates					
	Overall	Dawn	Noon	Dusk	Midnight	Average	All	Error	Dawn	Noon	Dusk	Midnight
Kp	5	5	5	5	5	5	7.26E+00	4.51E-01	6.16E+00	1.26E+01	5.10E+00	5.48E+00
PCI (average)	5	5	5	5	4	4.75	1.05E+02	8.30E+00	9.83E+01	1.97E+02	8.23E+01	5.91E+01
Intensity	5	5	5	5	3	4.5	4.17E-01	3.68E-02	4.39E-01	5.21E-01	5.28E-01	3.62E-01
PCI (stdev)	5	2	5	4	5	4	2.73E+02	2.64E+01	1.92E+02	4.88E+02	2.06E+02	2.08E+02
IMF B_z (stdev)	5	4	5	3	4	4	4.87E+01	4.45E+00	4.18E+01	6.78E+01	3.09E+01	3.93E+01
IMF B_y (stdev)	5	5	5	2	3	3.75	4.53E+01	4.22E+00	7.08E+01	5.46E+01	2.36E+01	2.98E+01
IMF B_x (stdev)	5	5	5	1	4	3.75	5.16E+01	6.00E+00	7.57E+01	6.20E+01	2.51E+01	4.57E+01
Season	5	5	1	2	5	3.25	1.62E+02	1.84E+01	2.20E+02	1.21E+02	1.01E+02	2.03E+02
Dst (average)	5	3	5	1	3	3	−5.05E+00	5.76E-01	−4.26E+00	−1.12E+01	−2.37E+00	−3.64E+00
IMF B_y (absolute average)	5	0	5	5	1	2.75	2.26E+01	2.84E+00		4.07E+01	2.56E+01	1.18E+01
HmF2	1	1	5	5	0	2.75	2.02E-04	9.76E-05	−4.32E-04	1.46E-03	8.02E-04	
HmE	5	0	5	3	0	2	1.35E-03	2.93E-04		3.58E-03	2.13E-03	
IMF B_z (average)	5	0	5	2	0	1.75	−1.21E+01	2.48E+00		−2.61E+01	−1.31E+01	
IMF B_y (average)	1	3	1	3	0	1.75	3.84E+00	1.86E+00	1.38E+01	8.03E+00	−1.08E+01	
Clock angle	5	0	5	2	0	1.75	−6.22E+01	1.23E+01		−1.36E+02	−6.62E+01	
E peak Ne	5	3	2	0	0	1.25	1.53E+01	3.38E+00	2.21E+01	2.95E+01		
*F*10.7 cm solar flux	1	0	3	1	0	1	3.20E-01	1.35E-01		1.11E+00	5.55E-01	

Note. Ranking and color conventions are as per Table 3. IMF = interplanetary magnetic field; PCI = polar cap index.

The importance of variations in the solar wind were indicated by the presence of explanatory variables $B_{Y(SD)}$ and $B_{Z(SD)}$. $B_{Z(SD)}$ is significant at all times of day, while $B_{Y(SD)}$ is only significant at noon and dusk. In all cases where these parameters are significant, they indicate that F_{Ne} increases as the variability in the solar wind increases. Variations in B_Y and B_Z can cause variations in the high-latitude convection pattern; they can also trigger particle precipitation.

The *Dst* index was a highly significant explanatory variable in the dawn, dusk, and midnight sectors, indicating that geomagnetic activity had a significant effect upon F_{Ne} at these locations. This was less significant than explanatory variables that indicated plasma transport and so suggested that precipitation was a secondary effect. It was surprising that *Dst* was more significant than *Kp*, as *Dst* is determined using measurements from a network of magnetometers at equatorial latitudes, whereas *Kp* is determined from a network of magnetometers at mid latitudes. The average ranking score for *Dst* was 2 (Table 3). The average ranking score for *Kp* was less than 1, and hence, it is not shown in Table 3. Variations in *Kp* are typically associated with both geomagnetic substorms and storms, whereas variations in *Dst* are primarily associated with ring currents. This suggests that it is the occurrence of geomagnetic storms, rather than substorms, that most strongly influences F_{Ne}.

The influence of the neutral atmosphere is seen both in the dusk and midnight sectors, with N_e increasing with the magnitude of U_{zon} in both of these sectors and increasing with the magnitude of U_{merid} in the midnight sector. This effect was less significant than plasma transport or precipitation but does indicate a link between the ionosphere and neutral atmosphere. There may also be a more tenuous link between the ionospheric and neutral atmosphere parameters. Vickers et al. (2014) showed that there is an approximately linear correlation between the density of atomic oxygen and the *F*10.7 cm solar flux. Hedin and Mayr (1987) showed a link between solar EUV output and thermosphere temperatures. Both the density of atomic oxygen and the temperature of the neutral atmosphere influence the density of the *F*-layer peak, through plasma production and recombination (Brekke, 1997).

A significant link can be observed between F_{Ne} and other ionospheric observations; F_{Ne} increases when the altitudes of the *E*- and *F*-layer peaks (hmE and hmF2) also increase. It also increases with increasing ion

Table 6
Neutral temperatures (T_n). Ranking and color conventions as per Table 3

	T_n											
	Significance						Parameter estimates					
Dependent Variable	Overall	Dawn	Noon	Dusk	Midnight	Average	All	Error	Dawn	Noon	Dusk	Midnight
Dst (average)	5	4	1	5	3	3.25	−4.70E-03	5.99E-04	−3.73E-03	−3.54E-03	−7.87E-03	−4.42E-03
Dst (stdev)	5	5	3	3	2	3.25	1.01E-03	1.32E-04	1.25E-03	1.21E-03	9.70E-04	6.55E-04
UT	0	3	5	2	0	2.5	−1.46E-02	8.41E-03	1.14E-01	−2.71E-01	−9.48E-02	
F10.7 cm solar flux	5	3	0	4	2	2.25	7.29E-04	1.36E-04	7.43E-04		1.04E-03	7.31E-04
Zenith wind	5	0	2	2	5	2.25	1.66E-04	3.06E-05		1.85E-04	1.74E-04	4.02E-04
F peak Ne	3	2	4	2	0	2	−1.33E-02	3.61E-03	−3.15E-02	−3.38E-02	−2.14E-02	
HmE	5	2	2	4	0	2	−1.82E-06	2.86E-07	−1.75E-06	−1.72E-06	−3.05E-06	
Zenith wind (magnitude)	2	0	2	1	5	2	9.70E-09	3.08E-09		2.49E-08	1.50E-08	3.70E-08
E region STR v1	5	1	3	1	2	1.75	−4.30E-02	7.86E-03	−3.78E-02	−3.78E-02	−5.99E-02	−9.02E-02
SSN	5	1	0	4	1	1.5	6.52E-04	1.37E-04	5.22E-04		1.10E-03	6.83E-04
HmF2	4	3	0	3	0	1.5	3.85E-07	8.91E-08	4.49E-07		6.28E-07	
PCI (average)	0	0	0	5	0	1.25	1.60E-02	8.27E-03			7.76E-02	
IMF B_z (Stdev)	4	1	0	0	3	1	1.69E-02	4.21E-03	1.48E-02			3.62E-02
IMF B_y (Stdev)	4	1	0	1	2	1	1.63E-02	3.77E-03	2.07E-02		1.85E-02	2.43E-02
Intensity	0	0	1	1	2	1	1.87E-05	3.23E-05		1.04E-04	2.08E-04	−3.16E-04

Note. IMF = interplanetary magnetic field; PCI = polar cap index.

temperature (T_i). Higher ion temperatures can indicate upwelling in the atmosphere, resulting in higher layer peaks. As the density of the atmosphere decreases with altitude so does plasma loss by recombination, leading to longer plasma lifetimes and hence densities.

5.2. Plasma Structuring in the *F* Region (F_{str})

The results from the single-term models for *F* region structuring are shown in Table 4. Season is clearly the most influential explanatory variable in the ranking order. It can be observed that each of the season coefficients for all time sectors are negative, indicating a that the strongest plasma density structures are observed closest to midwinter. These results are consistent with the seasonal dependencies reported by Wood and Pryse (2010) who suggested that this difference was primarily due to variation in the chemical composition of the atmosphere, which, in summer, both reduced the electron densities of the plasma drawn into the polar cap on the dayside and enhanced plasma loss by recombination. A secondary factor in this earlier study was the maintenance of the background polar ionosphere by photoionization in summer.

The seasonal influence may also be related to variations in the geometries of Earth's orbit. The elliptical nature of Earth's orbit results in a variation in the IMF field strength throughout the year, such that |IMF| is some 7% higher in January than in July (Newell et al., 2002). These authors also report that the southward component of the IMF maximizes in February thus implying more frequent reconnection events at these times. Particle precipitation that does occur at these times is therefore likely to involve greater populations, even in the absence of transient solar wind events. Plasma production by photoionization also varies throughout the year due to changes in the Sun-Earth distance, with plasma production ~6% higher in January than in July (Hargreaves, 1992).

Of lesser importance but still significant at rank 4.75 is the presence of E_{Ne}, which correlates positively with F_{str}, indicating a physical link which influences both the *F* region and the *E* region, simultaneously. The most likely candidate mechanism is particle precipitation; high energy particles entering the atmosphere at a steep angle stimulate ion production over a wide range of altitudes. The effect of this variable is greatest in the noon sector, and least at midnight, which may be due to changes in both the altitude and latitude at which precipitation is observed, relative to the location of the ESR. At midnight this region is largely equatorward of the ESR, whereas in the noon sector it is closer in latitude. Precipitating particles typically deposit

Table 7
Difference Between Ion and Neutral Temperatures ($T_i - T_n$)

$T_i - T_n$												
	Significance						Parameter estimates					
Dependent Variable	Overall	Dawn	Noon	Dusk	Midnight	Average	All	Error	Dawn	Noon	Dusk	Midnight
Intensity	5	5	5	2	5	4.25	3.96E-01	4.47E-02	4.88E-01	4.24E-01	3.57E-01	6.57E-01
Kp	5	3	5	4	2	3.5	6.26E+00	6.47E-01	4.81E+00	1.40E+01	5.26E+00	3.18E+00
PCI (average)	5	5	5	1	1	3	9.27E+01	1.16E+01	1.23E+02	2.19E+02	5.56E+01	3.87E+01
IMF B_z (stdev)	5	2	5	3	0	2.5	3.87E+01	5.99E+00	3.03E+01	6.77E+01	4.11E+01	
HmE	5	0	5	5	0	2.5	2.64E-03	4.09E-04		4.80E-03	4.95E-03	
T_e	5	5	0	5	0	2.5	1.85E-02	2.67E-03	1.67E-02		1.16E-01	
IMF B_y (Absolute average)	5	0	5	4	0	2.25	1.71E+01	3.80E+00		5.13E+01	2.70E+01	
HmF2	0	2	5	0	2	2.25	−6.30E-05	1.29E-04	−6.04E-04	1.78E-03		−6.63E-04
PCI (stdev)	5	2	5	1	0	2	2.25E+02	3.52E+01	2.19E+02	4.11E+02	1.55E+02	
IMF B_y (Stdev)	5	2	5	1	0	2	3.04E+01	5.37E+00	3.49E+01	4.44E+01	2.69E+01	
F peak Ne	2	0	3	4	0	1.75	1.39E+01	5.18E+00		4.48E+01	3.48E+01	
Dst (stdev)	5	2	0	1	3	1.5	−1.01E+00	1.93E-01	−1.42E+00		−7.01E-01	−1.07E+00
Season	5	2	0	0	4	1.5	1.49E+02	3.21E+01	1.72E+02			1.89E+02
E peak Ne	5	5	0	0	1	1.5	2.33E+01	4.26E+00	3.03E+01			2.06E+01
Zenith wind	3	0	1	0	5	1.5	−1.71E-01	4.42E-02		−2.21E-01		−4.81E-01
IMF B_x (stdev)	5	3	2	0	0	1.25	3.80E+01	7.89E+00	7.04E+01	4.26E+01		
Zonal wind (magnitude)	2	0	0	5	0	1.25	6.72E-04	2.30E-04			4.63E-03	
UT	1	0	5	0	0	1.25	2.93E+01	1.20E+01		3.05E+02		
E region STR v1	2	2	0	0	2	1	3.52E+01	1.14E+01	9.63E+01			1.18E+02
Zenith wind (magnitude)	0	0	0	0	4	1	−8.50E-06	4.40E-06				−3.84E-05

Note. Ranking and color conventions are as per Table 3. IMF = interplanetary magnetic field; PCI = polar cap index.

lower energies in the magnetic midnight sectors and deposit these energies at higher altitude (e.g., Vickrey et al., 1982), and so particle precipitation has less of an effect in the *E* region at midnight than at noon. The relatively high ranking of T_e in these results support the interpretation that particle precipitation is an important mechanism; electrons in the *F* region are energized by the passage of precipitating particles. The statistical significance of IMF B_Z in all time sectors, and *Kp* at noon and dawn, supports this interpretation, as both of these proxies are lined to particle precipitation.

The polar cap index (PCI: Troshichev et al. (1979), Troshichev and Andrezen (1985)) achieves a similarly high ranking. PCI is affected by solar wind velocity, southward IMF (giving rise to more frequent reconnection with Earth's magnetosphere), all of which contribute to the occurrence of substorm conditions and hence increased particle precipitation. Plasma transport across the polar cap from noon to midnight is also enhanced under these conditions. This is supported by the lower ranked but still influential IMF B_Z and clock angle, for which all statistically significant coefficients are negative, indicating that as the IMF rotates to a more southerly orientation, an increase in plasma structuring is observed.

Burns et al. (2004) suggested that variation in the neutral composition could also contribute to the formation of patches. In a modeling study neutral species were drawn antisunward with the TOI due to collisions. Interactions with neutrals already in the polar cap resulted in Joule heating, resulting in a pressure gradient that transported molecular rich air from polar cap to daytime midlatitudes. This process fragmented the TOI due to enhanced recombination, resulting in plasma structuring.

hmF2 positively influences the appearance of structuring in all time sectors except dawn. In general as altitude increases, so does structuring, particularly at dusk. At high altitude the more rarefied particle density reduces the collision and recombination frequency, enabling the structure to persist for a greater period of time and hence be transported over a greater distance. In the dawn time sector, the relationship between hmF2 and F_{str} reverses.

5.3. Ion Temperatures (T_i)

Table 5 shows that the *Kp* index bears the strongest statistical relationship with T_i in all time sectors and, based on the parameter estimates, has the largest effect in the noon sector, indicating the importance of cusp particle precipitation. As with the results for F_{str}, PCI features strongly in this analysis and repeats the same pattern, namely, that its influence is weakest in the midnight sector due to the extended distance over which ions from the dayside must traverse to reach this sector. The larger distance to terminator offers the greatest number of potential ion-neutral collisions by which the ions cool by shedding their thermal energy to the thermosphere. Support for this argument comes from the highly ranked influence of season (which acts as a proxy for the solar illumination angle), with lowest temperatures observed closest to midwinter when the distance to the terminator maximizes.

The variability of the IMF for all three field components ($B_{X(SD)}$, $B_{Y(SD)}$, and $B_{Z(SD)}$) and also PCI_{SD} are statistically significant, with increased values of T_i in all cases. Variations in the IMF cause variations in both the high-latitude convection pattern, which can cause flow channel events and elevate T_i. These variations in the IMF can also trigger substorms, which can elevate T_i through Joule heating from strong electric fields and particle precipitation.

The enhanced red line emission intensity (*I*), due to emission from atomic oxygen, is a significant explanatory variable. Solar illumination on the dayside both heats ions and excites neutrals, and these populations are transported over the terminator, where the neutral atoms deexcite, emitting photons. The influence of *I* upon T_i is weakest in the midnight sector, again due to the distance to the terminator.

At noon and dusk, T_i increases with increasing hmF2, and this is possibly related to upwelling of the atmosphere on the illuminated dayside, followed, once again, by cross-terminator transport processes. At dawn as hmF2 increases, Ti decreases. In the absence of a ionisation source, the plasma decays and cools. A higher peak altitude indicates longer-lived plasma as the plasma decays faster at the lower altitudes where recombination rates are higher. The relationship between hmF2 and Ti therefore suggests plasma long-lived plasma transported here from another location, possibly from the dayside.

5.4. Neutral Temperatures (T_n)

The most important explanatory variable for modeling T_n was *Dst*, and this can be seen in Table 6. All *Dst* coefficients are negative, indicating that as *Dst* moves to more extreme values (given that it has a negative value convention), the temperature of the neutral thermosphere increases. This effect is most pronounced in the dusk time sector. Higher *Dst* index values signify stronger equatorial current systems and a disturbed magnetosphere, typical of geomagnetic storms, which also affect high latitudes. Strong electric fields associated with particle precipitation at high latitudes are a well-known source of thermospheric Joule heating (Clausen et al., 2014; Foster et al., 1983; Knipp et al., 2004). Similarly, the presence of both U_{zen} and $|U_{zen}|$ at moderately high rankings can also be explained by corresponding induced Joule heating, which is known to generate strong advecting vertical winds above the region of heating due to upwelling (Chang & St.-Maurice, 1991).

The solar *F*10.7 cm flux is a general proxy for solar activity and related geomagnetic disturbances. The solar output at this wavelength is correlated with the EUV wavelengths that cause photoionization, which indirectly raises the temperature of the neutrals by way of increased Joule heating. Knipp et al. (2004) investigated the relative contributions of Joule heating, direct solar insolation, and particle precipitation on the variability of thermospheric temperatures during the period 1975–2003. These authors reported that, of the 595 GW of heating power input to the high-latitude thermosphere, 36 GW was attributed to particle precipitation, 95 GW was attributed to Joule heating, and the remainder to solar insolation. They also found that as solar activity (and hence *F*10.7 cm flux) increases, the variability of Joule heating also increases substantially.

The relatively high ranking of *F*10.7 and universal time in our analysis is also consistent with the findings of Hedin and Mayr (1987) who reported a correlation between variation in solar EUV output and variations in thermosphere temperatures. It may be that heated neutrals are present in this data set as a consequence of cross-terminator transport, which is supported by the observation that *F*10.7 has the weakest influence over T_n in the midnight and dawn sectors. The reasons for the absence of a strong positive coefficient for *F*10.7 in the noon sector are not clear.

T_n increases F_{Ne} decreases in all time sectors where it is statistically significant (dawn, noon, and dusk). This is a likely consequence of plasma recombination, the rate of which is enhanced when T_n increases, as shown in reaction rates given by McFarland et al. (1973), St.-Maurice and Torr (1978), and Hierl et al. (1997).

The reason for an absence of a strong positive coefficient for red line emission intensity when modeling T_n is not immediately clear. Several possibilities may account for this behaviour. All temperature measurements in this analysis are strongly dependent on altitude. The ESR data used here records T_i at an altitude of 240 km, to coincide with the assumed peak airglow emission based on the findings of Aruliah et al. (2005). Therefore, any substantially heated populations of ions at some altitude other than 240 km are not likely to be well observed in the data. If the peak airglow emission occurred at a higher, or lower, altitude then the ion temperature measurements would not be colocated. Plasma loss by recombination can also result in airglow.

5.5. Difference Between Ion and Neutral Temperatures ($T_i - T_n$)

The results for the single-term models of $T_i - T_n$ are shown in Table 7; clearly, the most statistically significant explanatory variable across all time sectors is intensity. The *Kp* index, PCI, and IMF B_Z are also highly ranked, and this is perhaps not surprising given their similarly high influence rankings for T_i discussed above. By contrast, the comparatively low ranking of variables, which may affect T_n such as U_{zen}, $|U_{zon}|$, or universal time, suggests that this relationship is dominated by variability in the ionosphere rather than the neutral atmosphere.

Burns et al. (2004) showed that neutral species are drawn antisunward with the TOI due to collisions. In this modeling study, interactions with neutrals already in the polar cap resulted in Joule heating. The transport of neutrals was less efficient than for ions, which may explain why explanatory variables associated with T_i rather than T_n dominate our models antisunward of the terminator.

6. Multiterm Model Results

Collectively the five dependent variables modeled in this study explain some of the behavior of the ionosphere-thermosphere system, such as the ion and neutral temperatures. The multiterm models show which *combination* of explanatory variables best explain the observed ionosphere-thermosphere variations observed in each case. This should not be interpreted as saying that any individual explanatory variable, which is rejected by our process is physically unimportant. The results in section 5 demonstrate numerous cases in which explanatory variables, which are rejected by the multiterm process are nevertheless important individually.

Previous studies have shown that ionospheric and thermospheric behavior can be influenced by preceding geomagnetic events on the order of hours (e.g., Aruliah et al., 1999). Finite times are also involved in the transport of energy from the impact of solar wind and IMF on the upstream magnetosphere, through flux transport to the magnetotail, reconnection processes, and ultimately, particle precipitation into the ionosphere. Several geophysical parameters used here, namely, IMF strength and variations in clock angle are provided from the Advanced Composition Explorer database already time lagged. However, appropriate time lags for model effective ionospheric and thermospheric behavior, such as changes in neutral wind velocity, are not as well known.

Single-term models were also produced for each dependent variable using neutral atmosphere parameters with time lags of up to ±6 hr. This process generated several hundred individual single-term models. Each time-lagged model was then cross correlated with the data for its dependent variable from the test data. All models for a given dependent variable were then ranked by their correlation, in descending order. All single-term models with a correlation of $<|0.2|$ were discarded. Of the remaining models, for each explanatory variable, the model with the highest correlation with the respective dependent variable was selected and used in place of its corresponding nonlagged term in the models in the following section. For example, if it were found that meridional wind velocity had a lower cross correlation with its dependent variable than, say, the negative 3-hr time-lagged meridional wind velocity, $U_{merid}(-3)$, then the unlagged term in the multiterm models would be supplanted with its lagged counterpart.

Physical interactions between the *F*-region ionosphere and the thermosphere can be demonstrated in instances where both ionospheric parameters, such as T_i and F_{Ne}, and thermospheric parameters, such as

T_n and I, are present in the final model. Models were produced for all five dependent variables listed in section 3.1, for all time sectors, plus for an all-day model. This process was repeated for all five dependent variables, using time-lagged supplanted terms. This enabled determination of whether the multiterm models produced were better at predicting the dependent variables in Table 1, when the most effective time-lagged parameters were included or without them. Both the time-lagged and unlagged models were tested for goodness of fit against the test data; the end results of this process are shown in Tables 8 and 9. In this study, only the multiterm models for T_n yielded a higher cross correlation with the T_n test data when using time-lagged terms, and this is also indicated in Table 8. Collectively, these results show that, for a predictive model, geophysical, neutral atmosphere and ionospheric terms need to be considered.

6.1. Goodness-of-Fit Testing

The goodness of fit of each of these models was established, using a number of goodness-of-fit statistics. The mean square error (MSE) measures the precision of the model. This statistic is heavily penalized by extreme values. The relative mean error (ME) measures bias in the model. It is given by the ratio of the mean of the residuals to the mean of the observations. Morley et al. (2018) introduced the *median symmetric accuracy* (MSA), which penalizes all values in the same way, regardless of magnitude. Given that the MSA yields a percentage value, it can be easily interpreted across all models given here, irrespective of differences in the numerical magnitudes involved, or how data distributions are expressed. It is defined as

$$MSA = 100\left(\exp\left(M\left(\left|\log_e\left(\frac{y}{x}\right)\right|\right)\right)-1\right), \tag{5}$$

where M is the median function and y/x is the accuracy ratio where y is the predicted value and x is the observed value (Tofallis, 2015).

In order to establish whether our models performed better than climatology, the skill scores were computed as per equation (6):

$$Score = 1-\left(\frac{\sigma_{\text{Model}}}{\sigma_{\text{Climate}}}\right), \tag{6}$$

where σ_{Model} and σ_{Climate} are the errors in the model and in the climatology, respectively. The MSE and MSA can be used interchangeably in equation (6) to produce two skill scores for each goodness-of-fit statistic. A skill score test, which yields a negative number indicates that the predictive error of the baseline climatology is smaller than the predictive error from the models. Skill score results that are ~0 indicate that both the models and the climatology are performing approximately as well as each other. A positive value indicates that the model is performing better than the climatology. All the goodness-of-fit statistics (ME, MSE, and MSA), the cross correlations between the models and the test data, and the skill scores for both MSA and MSE are shown in Table 9.

7. Discussion

This paper demonstrates the effectiveness of linear modeling as a technique for investigating various aspects of the behavior of the coupled ionosphere-thermosphere system, under polar darkness conditions. It demonstrates how such an approach offers more accurate predictions of plasma densities, structures, ion, and neutral temperatures than by climatology alone. The positive MSE skill scores demonstrate that the multiterm models are capturing more of the extreme behavior of the dependent variables than by the climatological approach. The MSA skill scores for most of the dependent variables are close to 0, indicating that the improvement in predicting extreme values has not come at the expense of the more common, less extreme, values, which comprises the bulk of the data set. The exception to this is F_{str} where the MSA skill scores take positive values between 0.13 and 0.50. This demonstrates that the models are better representing observations across the entire data set. Finally, for all dependent variables, the small values of the relative ME show that the models are not heavily subject to bias.

To demonstrate this point, Table 9 shows the variation in skill scores through different time sectors, based on an MSE goodness-of-fit test; in all cases the skill score is positive, showing that the multiterm models are

Journal of Geophysical Research: Space Physics

10.1029/2018JA026171

Table 8
The Final Results of the Multiterm Modeling Process Showing the Combinations of Explanatory Variables That Predict Ionosphere-Thermosphere Behavior

	All time sectors	Dawn	Noon	Dusk	Midnight
T_i (unlagged)	$T_i = 683.24 + (9.08F_{str}) + (0.75T_e) + (0.47F_{10.7}) + (27.97B_{Z(SD)}) + (204.27doy)$	$T_i = 387.36 + (47.1E_{Ne}) + (0.83T_e) + (4Kp) + (12.1B_Y)$	$T_i = 364.24 + (0.61T_e) + (1.11U_{mer}) - (144.9C) + (150B_{Z(SD)})$	$T_i = 935.05 + (1.04U_{zon}) + (1.85F_{10.7}) + (8.71Kp) - (37.7B_{X(SD)}) + (279.05doy)$	$T_i = 603.5 + (12.9F_{str}) + (0.51T_e) + (0.59I) + (22.3B_{Y(SD)}) + (15.2B_Z) + (155doy)$
T_n (lagged)	$\log(T_n) = 6.73 + (1.51\times10^{-4}T_e) + (10^{-3}U_{mer}(-3)) + (0.011doy)$	$\log(T_n) = 2.93 + (8.04\times10^{-5}) + (3\times10^{-4}U_{mer}(-3)) + (8.7\times10^{-2}doy)$	$\log(T_n) = 6.81 + (2.59\times10^{-4}U_{zen}(0.5)) + (1.45\times10^{-3}F_{10.7}) - (0.21C_{sd})$	$\log(T_n) = 6.85 - (2.5\times10^{-2}F_{Ne}) - (3.04\times10^{-6}hmE) + (1.69\times10^{-3}F_{10.7}) + (4.2\times10^{-3}Kp)$	$\log(T_n) = 6.71 + (2.62B_{Y(SD)}) + (10^{-3}U_{mer}(-3)) - (1.3\times10^{-3}U_{zon})$
$T_i - T_n$ (unlagged)	$T_i - T_n = -352.2 + (0.45T_e) + (0.62I) - (0.43U_{mer}) + (0.85U_{zon}) + (171.6C_{sd})$	$T_i - T_n = (55.4F_{Ne}) + (34.7F_{str}) + (0.26T_e)$	$T_i - T_n = 294 - (0.48I) - (9.5\times10^{-4}hmF2) + (3.17U_{mer}) + (8.02Kp)$	$T_i - T_n = 300.8 + (0.61I) + (0.85U_{mer}) + (1.29U_{zon})$	$T_i - T_n = -318.9 + (0.46T_e) + (0.98I)$
F_{Ne} (unlagged)	$\log(F_{Ne}) = -1.92 + (2.07\times10^{-6}hmF2) + (9.85\times10^{-6}hmE) + (8\times10^{-4}I) + (0.43E_{str})$	$\log(F_{Ne}) = -0.58 + (0.47E_{str}) - (5.9\times10^{-4}T_e) + (9\times10^{-3}F_{10.7})$	$\log(F_{Ne}) = -0.81 + (0.92E_{str}) + (10^{-2}F_{10.7})$	$\log(F_{Ne}) = -0.4 + (0.35E_{str}) + (9.6\times10^{-3}F_{10.7}) + (0.13B_{Y(SD)}) + (0.89doy)$	$\log(F_{Ne}) = -1.52 + (0.62E_{str}) + (1.4\times10^{-2}F_{10.7})$
F_{str} (unlagged)	$\log(F_{str}) = 0.27 + (0.22E_{Ne}) + (9.24\times10^{-5}T_i) + (10^{-2}C_{sd})$	$\log_e(F_{str}) = 0.33 + (0.14E_{Ne}) - (0.13C) + (5.75\times10^{-2}B_{Y(SD)})$	$\log(F_{str}) = -0.28 + (0.47E_{Ne})$	$\log(F_{str}) = -0.48 - (0.2doy) (1.2\times10^{-6}hmF2) + (0.31E_{Ne}) - (1.8\times10^{-4}T_i)$	$\log(F_{str}) = 0.58 + (0.58E_{str}) - (1.4\times10^{-3}I) - (0.56doy)$

Note. Cross correlations with the test data and goodness-of-fit statistics are shown for each of these models in Table 9. Only the model for T_n performed better in prediction tests when using time-lagged terms. In all other instances, the lagged multiterm models either performed no better or were worse than the unlagged models.

Table 9
The Goodness-of-Fit Statistics Calculated for Each of the Time Sector Multiterm Model (Shown in Table 8)

MIDNIGHT	Model ME	Model MSA	Model MSE	Correlation	MSA skill score	MSE skill score
T_i	0.02	39.7	19619	0.76	0.04	0.58
$\log_e(T_n)$	0.006	37.4	0.029	0.36	0.001	0.14
$T_i - T_n$	−0.01	77.0	44851	0.59	0.001	0.35
$\log_e(F_{Ne})$	−0.136	75.0	0.369	0.75	−0.03	0.77
$\log_e(F_{str})$	0.05	51.8	0.211	0.40	0.13	0.21
DAWN						
T_i	0.04	39.7	37332	0.59	0.03	0.49
$\log_e(T_n)$	0.018	37.1	0.024	0.39	0.01	0.84
$T_i - T_n$	0.17	67.1	49004	0.50	0.08	0.32
$\log_e(F_{Ne})$	0.06	84.7	0.236	0.72	0.03	0.51
$\log_e(F_{str})$	0.039	47.7	0.072	0.78	0.16	0.61
NOON						
T_i	0.02	41.9	37836	0.57	−0.02	0.32
$\log_e(T_n)$	−0.003	37.3	0.021	0.54	0.01	0.29
$T_i - T_n$	−0.04	76.1	40883	0.46	−0.04	0.44
$\log_e(F_{Ne})$	−0.07	65.2	0.5126	0.61	−0.02	0.36
$\log_e(F_{str})$	0.003	39.0	0.0304	0.92	0.50	0.88
DUSK						
T_i	−0.01	39.7	27994	0.43	0.02	0.25
$\log_e(T_n)$	−0.0005	37.3	0.019	0.50	0.002	0.25
$T_i - T_n$	−0.09	60.3	18191	0.34	0.04	0.50
$\log_e(F_{Ne})$	0.05	64.9	0.33	0.69	0.14	0.47
$\log_e(F_{str})$	0.042	41.4	0.04	0.87	0.30	0.80

Note. The cross correlation between the model and the test data for each dependent variable is also shown, as are the results of the MSA and MSE skill score tests performed. In all cases where MSE is used as the goodness-of-fit, positive values are found indicating that the multiterm model is performing significantly better in predicting the behavior of the coupled ionosphere-thermosphere than the climatology. The difference in magnitudes for the MSE for both T_i and $T_i - T_n$ is because these variables were modeled with a normal distribution, unlike the other variables. Although some of the model MSE errors look large, these are reasonable. For example the MSE of 19619 for T_i in the midnight sector corresponds to a value of 140 K when the square root is taken, this is an order of magnitude smaller than the typical values for T_i. Under the MSA convention the skill scores are either close to 0, indicating no substantial difference in efficacy of each method, or positive, indicating, again, that the models are outperforming the climatology. MSA = median symmetric accuracy; MSE = mean square error.

outperforming climatology. Climatology has traditionally been used to represent the averaged behavior of the ionosphere. Here we demonstrate one approach, which improves upon climatology.

Table 9 also shows that statistically significant correlations exist between the multiterm models and the test data for all dependent variables, including $T_i - T_n$ with values ranging from 0.34 to 0.92. The skill score for the peak electron count in the *F*-region ionosphere (F_{Ne}) appears to show the greatest improvement in predictability with the model over climatology in the midnight sector. *F*-region structuring (F_{str}) performs best at noon and dusk. The Joule heating variable $T_i - T_n$ is most predictable in the dusk time sector and less so at other times. The model for T_n appears to reach a substantial maximum improvement in performance versus climatology at dawn.

The utility of time-lagged dependent variables was investigated to understand whether their inclusion in the multiterm models improved the predictability of ionosphere-thermosphere variability. This investigation was conducted for all five dependent variables; only in the case of neutral temperatures (T_n) was any measurable improvement in correlation between the model and the test data observed. In all other cases, time lagging offered no substantial improvement in predictability or was worse. For T_n, the explanatory variables that improved the predictability of the model were the meridional and zonal wind velocities, lagged at −3 and +0.5 hr, respectively (see T_n terms in Table 8). This is consistent with previous findings of Killeen et al. (1984) and Aruliah et al. (2004), who showed that nighttime response times for the thermospheric winds can be up to 3–6 hr.

An investigation of this kind was greatly simplified by the use of ground based solar-terrestrial physics facilities. These sample the same location throughout the study and are not subject to the changes in viewing geometry commonly encountered in space-based instruments. The ESR and the FPI are approximately colocated and have a substantial archive of data, enabling this type of long-term study.

The mixture of variables used in this study can be split broadly into three categories, based upon their physical interaction with whichever dependent variable is being modeled, these are causative, coincidental, and consequential variables. Causative variables have a direct physical influence over the behavior of a given dependent variable. Season, for example, is a causative variable, which strongly influences the behavior of F_{str}. E_{Ne} is a coincidental variable when associated with F_{str}; particle precipitation affects the upper atmosphere over a wide range of altitudes, producing plasma structuring in the F region, and simultaneously increasing electron density in the E region. The red line emission intensity is a consequential variable, which is influenced by the presence of heated ions in the F region (T_i). In all cases, however, these variables can still be used singularly or in combinations to predict some of the variability we observe in the ionosphere-thermosphere system. In this study, all three types of variables have been used to create a predictive model. For the purposes of producing operational models in future work, models will be further developed only using causative variables.

The models produced here are predictive in the sense that one may use the variables included, both singularly, and in the combinations shown, to predict to some degree the variability of the coupled ionosphere-thermosphere. In order for a model to be operational, one needs to observe some parameter combinations from independent data sources which, together, still enable predictability. For example, the multiterm model for T_i in the dawn sector (Table 8) includes the terms for T_e, and E_{Ne} both of which are measured from the same data source as T_i, namely, the ESR. However, the multiterm model for T_i in the dusk sector, contains only terms that are recorded from sources other than the ESR, and hence, this is an example of an operational model. Improving the operational facility of these models is a goal for future work in this area. The reason why an operational model relies on the availability of a data source, which is independent of the modeled parameter itself, is because one could simply measure the parameter directly with the same instrument, thus forgoing the task of modeling it to begin with. If, then, the ultimate goal of using models as predictive tools is to be realized, it will often be necessary, on a practical basis, to use operational models, as direct means of sampling a given parameter routinely may not always be available. If we can infer, say, T_e *reliably* from operational models which use data sources that are independent of T_e, then it would be possible to characterize T_e, without directly sampling it.

In all cases, except for $T_i - T_n$, geophysical parameters have the highest ranking for statistical significance in the single-term analysis. The generally lower rankings for T_n and the somewhat lower levels of correlation between the multiterm models for T_n, and the test data, imply that predicting T_n remains the dependent variable in this study with the most unpredictability. Certainly any improvement in predictive modeling for the thermospheric temperatures would substantially improve the related predictability of $T_i - T_n$. Vertical coupling within the atmosphere can affect thermospheric parameters (Yiğit et al., 2016), for example, during a Sudden Stratospheric Warming, T_n has been observed to vary at thermospheric altitudes (Yamazaki et al., 2015). Inclusion of such processes is likely to yield an improvement in the prediction of T_n.

Given the difficulties inherent in acquiring neutral thermosphere parameters with the FPI during daylight, to date, we have only applied this approach during polar darkness. Studies are underway, however, to model ionospheric parameters, including F_{Ne}, F_{str}, and ionospheric phase and amplitude scintillation, under all universal times throughout the year. These ongoing studies use data from a variety of sources, including GNSS receivers, the ESR, and ground magnetometers, as well as geohysical proxies.

8. Conclusions

In this study, we have demonstrated that statistical modeling can be used as a technique for predicting variability in the coupled high-latitude F-region ionosphere-thermosphere system, outperforming climatology. We have quantified the level of influence that ionospheric, neutral atmosphere, and geophysical parameters have on five key dependent variables: ion temperature, neutral temperature, plasma structuring, electron

density, and $T_i - T_n$. The models reveal both the individual parameters which, by themselves, offer the most significant explanatory power for ionosphere-thermosphere variability and which combinations of parameters offer likewise. The dominant geophysical proxy for the electron density at the *F*-layer peak was the *F*10.7 cm solar flux. For the variations in plasma density, it was the seasonal influence. In the cases of ion and neutral temperatures the dominant parameters were the *Kp* and *Dst* indices, respectively. For $T_i - T_n$ the intensity of airglow emission was found to be most important. These results also demonstrated that ionospheric variability tends to dominate the heat transfer relationship between the ionosphere and neutral atmosphere. This can be seen in the sense that the same high-ranked terms which, in the single-term models, predict T_i, also appear in the single-term results for $T_i - T_n$. The models also show which combinations of ionospheric and neutral atmosphere parameters together best explain observed variability. The skill scores for these models show an improvement upon climatology in every case, with values for skill scores based on the MSE of up to 0.88.

Acknowledgments
EISCAT is an international facility supported by China, Finland, Japan, Norway, Sweden, and the United Kingdom. The assistance of Steve Crothers and colleagues in the UK EISCAT Support Group at the Rutherford Appleton Laboratory is gratefully acknowledged. The IMF data were provided by N. Ness and obtained from the CDAWeb. The geophysical proxies were obtained from the U.K. Solar System Data Centre at the Rutherford Appleton Laboratory. The assistance of Myron Huzan with the preliminary data processing is gratefully acknowledged. The continual operation of the Svalbard FPI relies on Ian McWhirter, and we would also like to thank Eoghan Griffin and Andy Charalambous and the site managers at the Kjell Henriksen Observatory.

References

Aruliah, A. L., Griffin, E. M., Aylward, A. D., Ford, E. A. K., Kosch, M. J., Davis, C. J., et al. (2005). First direct evidence of meso-scale variability on ion-neutral dynamics using co-located tristatic FPIs and EISCAT radar in Northern Scandinavia, EISCAT workshop special issue. *Annales Geophysicae*, *23*, 147–162.

Aruliah, A. L., Griffin, E. M., McWhirter, I., Aylward, A. D., Ford, E. A. K., & Charalambous, A. (2004). First tristatic studies of meso-scale ion-neutral dynamics and energetics in the high-latitude upper atmosphere using collocated FPIs and EISCAT radar. *Geophysical Research Letters*, *31*, L03802. https://doi.org/10.1029/2003GL018469

Aruliah, A. L., Müller-Wodarg, I. C. F., & Schoendorf, J. (1999). Consequences of geomagnetic history on the high-latitude thermosphere and ionosphere: Averages. *Journal of Geophysical Research*, *104*(A12), 28,073–28,088. https://doi.org/10.1029/1999JA900334

Banks, P. M., & Kockarts, G. (1973). *Aeronomy*. New York: Academic Press.

Brekke, A. (1997). *Physics of the upper polar atmosphere*, *Wiley-Praxis Series in Atmospheric Physics*. Chichester.

Brekke, A., & Kamide, Y. (1996). On the relationship between Joule and frictional heating in the polar ionosphere. *Journal of Atmospheric and Terrestrial Physics*, *58*(1–4), 139–143. https://doi.org/10.1016/0021-9169(95)00025-9

Burns, A. G., Wang, W., Killeen, T. L., & Solomon, S. C. (2004). A "tongue" of neutral composition. *Journal of Atmospheric and Solar-Terrestrial Physics*, *66*(15-16), 1457–1468. https://doi.org/10.1016/j.jastp.2004.04.009

Burston, R., Astin, I., Mitchell, C., Alfonsi, L., Pederseon, T., & Skone, S. (2010). Turbulent times in the northern polar ionosphere? *Journal of Geophysical Research*, *115*, A04310. https://doi.org/10.1029/2009JA014813

Chang, C. A., & St.-Maurice, J.-P. (1991). Two-dimensional high-latitude thermospheric modeling—A comparison between moderate and extremely disturbed conditions. *Canadian Journal of Physics*, *69*, 1007–1031. https://doi.org/10.1139/p91-159

Chapman, S. (1931). The absorption and dissociative or ionizing effect of monochromatic radiation in an atmosphere on a rotating Earth Part II. Grazing incidence. *Proceedings of the Physical Society*, *43*(5), 483–501. https://doi.org/10.1088/0959-5309/43/5/302

Clausen, L. B. N., Milan, S. E., & Grocott, A. (2014). Thermospheric density perturbations in response to substorms. *Journal of Geophysical Research: Space Physics*, *119*, 4441–4455. https://doi.org/10.1002/2014JA019837

Cole, K. D. (1962). Joule heating of the upper atmosphere. *Australian Journal of Physics*, *15*(2), 223. https://doi.org/10.1071/PH620223

Crowley, G. (1996). Critical review of ionospheric patches and blobs. In W. R. Stone (Ed.), *Review of Radio Science 1993–1996* (Chap. 27, pp. 619–648). UK: Oxford Science Publication.

Dungey, J. W. (1961). Interplanetary magnetic field and the auroral zones. *Physical Review Letters*, *6*(2), 47–48. https://doi.org/10.1103/PhysRevLett.6.47

Elmas, Z. G., Aquino, M., Marques, H. A., & Monico, J. F. G. (2011). Higher order ionospheric effects in GNSS positioning in the European region. *Annales Geophysicae*, *29*, 1383–1399. https://doi.org/10.5194/angeo-29-1383-2011

Foster, J. C. (1984). Ionospheric signatures of magnetospheric convection. *Journal of Geophysical Research*, *89*(A2), 855–865.

Foster, J. C., St.-Maurice, J.-P., & Abreu, V. J. (1983). Joule heating at high latitudes. *Journal of Geophysical Research*, *88*(A6), 4885–4897. https://doi.org/10.1029/JA088iA06p04885

Frey, H. U., Phan, T. D., Fuselier, S. A., & Mende, S. B. (2003). Continuous magnetic reconnection at Earth's magnetopause. *Nature*, *426*(6966), 533–537. https://doi.org/10.1038/nature02084

Gonzalez, W. D., Joselyn, J. A., Kamide, Y., Kroehl, H. W., Rostoker, G., Tsuratani, B. T., & Vasyliunas, V. M. (1994). What is a geomagnetic storm? *Journal of Geophysical Research*, *99*(A4), 5771–5792. https://doi.org/10.1029/93JA02867

Hapgood, M. (2017). Satellite navigation—Amazing technology but insidious risk: Why everyone needs to understand space weather. *Space Weather*, *15*, 545–548. https://doi.org/10.1002/2017SW001638

Hargreaves, J. K. (1992). *The solar-terrestrial environment*, *Cambridge Atmospheric and Space Science Series* (Vol. 5). Cambridge, UK: Cambridge University Press.

Hays, P. B., Rusch, D. W., Roble, R. G., & Walker, J. C. G. (1978). The O I (6300 Å) airglow. *Reviews of Geophysics and Space Physics*, *16*(2), 225–232. https://doi.org/10.1029/RG016i002p00225

Hedin, A. E., & Mayr, H. G. (1987). Solar EUV induced variations in the thermosphere. *Journal of Geophysical Research*, *92*(D1), 869–875. https://doi.org/10.1029/JD092iD01p00869

Hierl, P. M., Dotan, I., Seeley, J. V., van Doren, J. M., Morris, R. A., & Viggiano, A. A. (1997). Rate constants for the reactions of O^+ with N_2 and O_2 as a function of temperature (300–1800K). *Journal of Chemical Physics*, *106*(9), 3540–3544. https://doi.org/10.1063/1.473450

Hubert, B., Milan, S. E., Grocott, A., Blockx, C., Cowley, S. W. H., & Gérard, J.-C. (2006). Dayside and night side reconnection rates inferred from IMAGE FUV and Super Dual Auroral Radar Network data. *Journal of Geophysical Research*, *111*, A03217. https://doi.org/10.1029/2005JA011140

Killeen, T. L., Hays, P. B., Carignan, G. R., Heelis, R. A., Hanson, W. B., Spencer, N. W., & Brace, L. H. (1984). Ion-neutral coupling in the high latitude *F*-region: Evaluation of ion heating terms from Dynamics Explorer 2. *Journal of Geophysical Research*, *89*(A9), 7495–7508. https://doi.org/10.1029/JA089iA09p07495

Kinrade, J., Mitchell, C. N., Smith, N. D., Ebihara, Y., Weatherwax, A. T., & Bust, B. S. (2013). GPS phase scintillation associated with optical auroral emissions: First statistical results from the geographic South Pole. *Journal of Geophysical Research: Space Physics*, *118*, 2490–2502. https://doi.org/10.1002/jgra.50214

Kintner, P. M., Ledvina, B. M., & de Paula, E. R. (2007). GPS and ionospheric scintillations. *Space Weather*, *5*, S09003. https://doi.org/10.1029/2006SW000260

Knipp, D. J., Tobiska, W. K., & Emery, B. A. (2004). Direct and indirect thermospheric heating sources for solar cycles 21–23. *Solar Physics*, *224*, 495. https://doi.org/10.1007/s11207-005-6393-4

Lorentzen, D. A., Shumilov, N., & Moen, J. (2004). Drifting airglow patches in relation to tail reconnection. *Geophysical Research Letters*, *31*, L02806. https://doi.org/10.1029/2003GL017785

McCullagh, P., & Nelder, J. A. (1983). *Generalized linear models*, *Monographs on Statistics and Applied Probability* (Vol. 37). New York: Chapman and Hall.

McEwen, D. J., & Harris, D. P. (1996). Occurrence patterns of F layer patches over the north magnetic pole. *Radio Science*, *31*(3), 619–628.

McFarland, M., Albritton, D. L., Fehsenfeld, F. C., Ferguson, E. E., & Schmeltekopf, A. L. (1973). Flow-drift technique for ion mobility and ion-molecule reaction rate constant measurements, II, positive ion reactions of N^+, O^+ and $N_2{}^+$ with O_2 and O_+ and N_2 from thermal to −2eV. *Journal of Chemical Physics*, *59*(12), 6620–6628. https://doi.org/10.1063/1.1680042

Millward, G. H., Moffett, R. J., Balmforth, H. F., & Rodger, A. S. (1999). Modeling the ionospheric effects of ion and electron precipitation in the cusp. *Journal of Geophysical Research*, *104*, 24,603–24,612.

Mitchell, C. N., Alfonsi, L., De Franceschi, G., Lester, M., Romano, V., & Wernik, A. W. (2005). GPS TEC and scintillation measurements from the polar ionosphere during the October 2003 storm. *Geophysical Research Letters*, *32*, L12S03. https://doi.org/10.1029/2004GL021644

Morley, S. K., Brito, T. V., & Welling, D. T. (2018). Measures of model performance based on the log accuracy ratio. *Space Weather*, *16*, 69–88. https://doi.org/10.1002/2017SW001669

Newell, P. T., Sotirelis, T., Skura, J. P., Meng, C. I., & Lyatsky, W. (2002). Ultraviolet insolation drives seasonal and diurnal space weather variations. *Journal of Geophysical Research*, *107*(A10), 1305. https://doi.org/10.1029/2001JA000296

Pedersen, T. R., Fejer, B. G., Doe, R. A., & Weber, E. J. (1998). Incoherent scatter radar observations of horizontal F region plasma structure over Sondrestrom, Greenland, during polar cap patch events. *Radio Science*, *33*(6), 1847–1866.

Pedersen, T. R., Fejer, B. G., Doe, R. A., & Weber, E. J. (2000). An incoherent scatter radar technique for determining two-dimensional horizontal ionization structure in polar cap F region patches. *Journal of Geophysical Research*, *105*, 10,637–10,655.

Pryse, S. E., Kersley, L., Malan, D., & Bishop, G. J. (2006). Parameterization of the main ionospheric trough in the European sector. *Radio Science*, *41*, RS5S14. https://doi.org/10.1029/2005RS003364

Rishbeth, H., & Setty, C. S. G. K. (1961). The F-layer at sunrise. *Journal of Atmospheric and Terrestrial Physics*, *21*, 263–276.

Rishbeth, H., & Williams, P. J. S. (1985). The EISCAT ionospheric radar: The system and its early results. *The Quarterly Journal of the Royal Astronomical Society*, *26*, 478–512.

Schunk, R. W. (1975). Transport equations for aeronomy. *Planetary and Space Science*, *23*(3), 437–485. https://doi.org/10.1016/0032-0633(75)90118-X

Sojka, J. J., & Schunk, R. W. (1987). Theoretical study of the high-latitude ionosphere's response to multicell convection patterns. *Journal of Geophysical Research*, *92*(A8), 8733–8744. https://doi.org/10.1029/JA092iA08p08733

Sojka, J. J., Bowline, M. D., & Schunk, R. W. (1994). Patches in the polar ionosphere: UT and seasonal dependence. *Journal of Geophysical Research*, *99*(A8), 14,959–14,970. https://doi.org/10.1029/93JA03327

Spogli, L., Alfonsi, L., De Franceschi, G., Romano, V., Aquino, M. H. O., & Dodson, A. (2009). Climatology of GPS ionospheric scintillations over high and mid-latitude European regions. *Annales Geophysicae*, *27*, 3429–3437. https://doi.org/10.5194/angeo-27-3429-2009

St.-Maurice, J. -P., & Torr, D. G. (1978). Nonthermal rate coefficients in the ionosphere: The reactions of O^+ with N_2, O_2 and NO. *Journal of Geophysical Research*, *83*(A3), 969–977. https://doi.org/10.1029/JA083iA03p00969

Tofallis, C. (2015). A better measure of relative prediction accuracy. *Journal of the Operational Research Society*, *66*(8), 1352–1362. https://doi.org/10.1057/jors.2014.103

Troshichev, O. A., & Andrezen, V. G. (1985). The relationship between interplanetary quantities and magnetic activity in the southern polar cap. *Planetary and Space Science*, *33*(4), 415–419. https://doi.org/10.1016/0032-0633(85)90086-8

Troshichev, O. A., Dmitrieva, N. P., & Kuznetsov, B. M. (1979). Polar cap magnetic activity as a signature of substorm development. *Planetary and Space Science*, *27*(3), 217–221. https://doi.org/10.1016/0032-0633(79)90063-1

Tsunoda, R. T. (1988). High-latitude F-region irregularities: A review and synthesis. *Reviews of Geophysics*, *26*, 719.

Vasyliunas, V. M., & Song, P. (2005). Meaning of ionospheric Joule heating. *Journal of Geophysical Research*, *110*, A02301. https://doi.org/10.1029/2004JA010615

Velinov, P. (1968). On ionization in the ionospheric D-region by galactic and solar cosmic rays. *Journal of Atmospheric and Terrestrial Physics*, *30*(11), 1891–1905. https://doi.org/10.1016/0021-9169(68)90031-7

Vickers, H., Kosch, M. J., Sutton, E., Bjoland, L., Ogawa, Y., & La Hoz, C. (2014). A solar cycle of upper thermosphere density observations from the EISCAT Svalbard radar. *Journal of Geophysical Research: Space Physics*, *119*, https://doi.org/10.1002/2014JA019885, 6833–6845.

Vickrey, J. F., Vondrak, R. R., & Matthews, S. J. (1982). Energy deposition by precipitating particles and Joule dissipation in the auroral ionosphere. *Journal of Geophysical Research*, *87*(A7), 5184–5196. https://doi.org/10.1029/JA087iA07p05184

Wannberg, G., Wolf, I., Vanhainen, L. G., Koskenniemi, K., Röttger, J., Postila, M., et al. (1997). The EISCAT Svalbard radar: A case study in modern incoherent scatter radar system design. *Radio Science*, *32*(6), 2283–2307. https://doi.org/10.1029/97RS01803

Weber, E. J., Klobuchar, J. A., Buchau, J., Carlson, H. C. Jr., Livingston, R. C., de laBeaujardiere, O., et al. (1986). Polar cap *F* layer patches: Structure and dynamics. *Journal of Geophysical Research*, *91*(A11), 12,121–12,129. https://doi.org/10.1029/JA091iA11p12121

Wood, A. G., & Pryse, S. E. (2010). Seasonal influence on polar cap patches in the high-latitude night side ionosphere. *Journal of Geophysical Research*, *115*, A07311. https://doi.org/10.1029/2009JA014985

Wright, J. W. (1963). The F region seasonal anomaly. *Journal of Geophysical Research*, *68*, 4379–4381.

Yamazaki, Y., Kosch, M. J., & Emmert, J. T. (2015). Evidence for stratospheric sudden warming effects on the upper thermosphere derived from satellite orbital decay data during 1967–2013. *Geophysical Research Letters*, *42*, 6180–6188. https://doi.org/10.1002/2015GL065395

Yiğit, E., Knížová, P. K., Georgieva, K., & Ward, W. (2016). A review of vertical coupling in the atmosphere–ionosphere system: Effects of waves, sudden stratospheric warmings, space weather, and of solar activity. *Journal of Atmospheric and Solar-Terrestrial Physics*, *141*, 1–12.

Zwickl, R. D., Doggett, K. A., Sahm, S., Barrett, W. P., Grubb, R. N., Detman, T. R., et al. (1998). The NOAA Real-Time Solar Wind (RTSW) System Using ACE Data. *Space Science Reviews*, *86*, 633–648.