Introduction to the Ionosphere

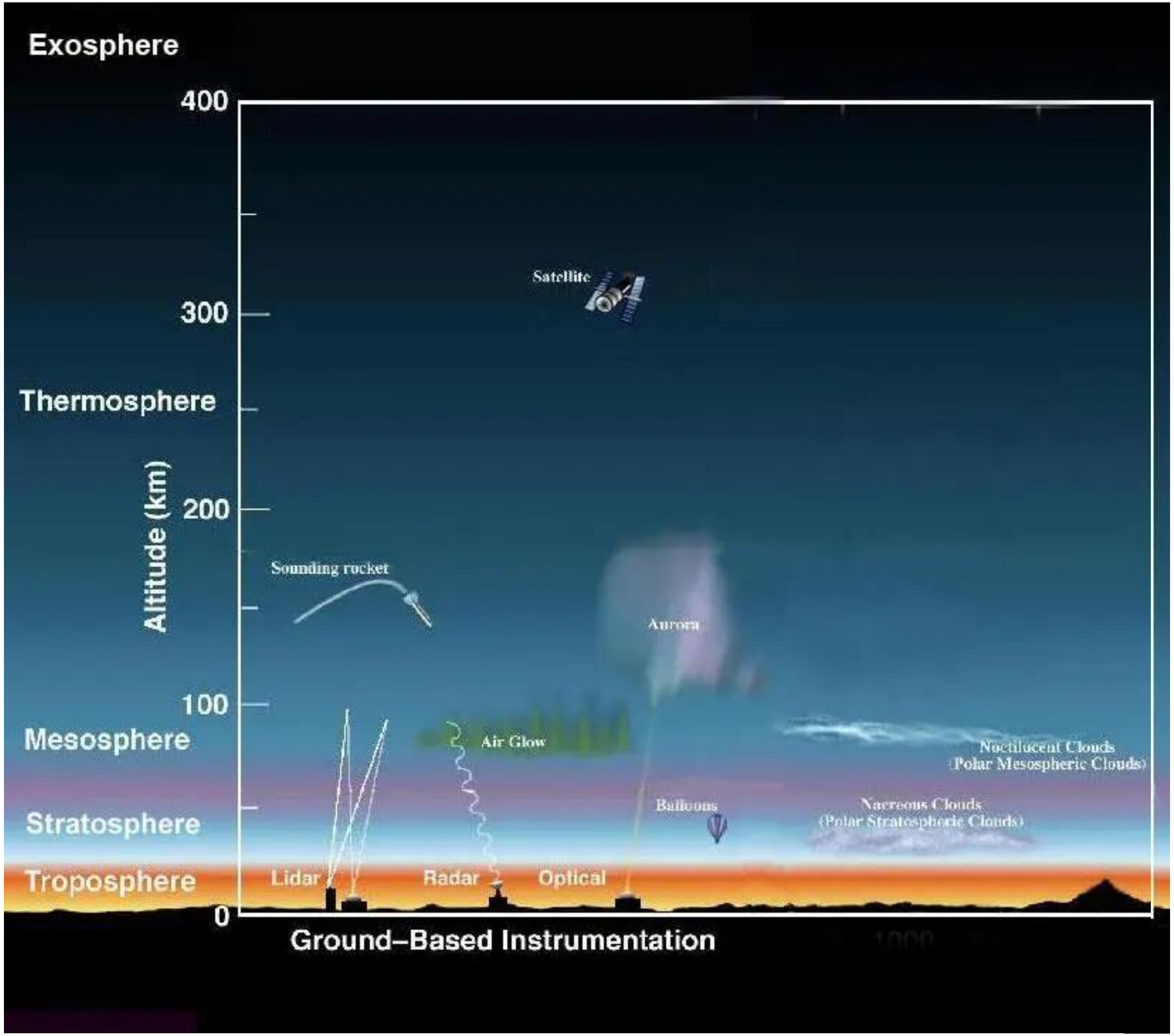
Andrew J. Kavanagh

Space Weather and Atmosphere Team
British Antarctic Survey

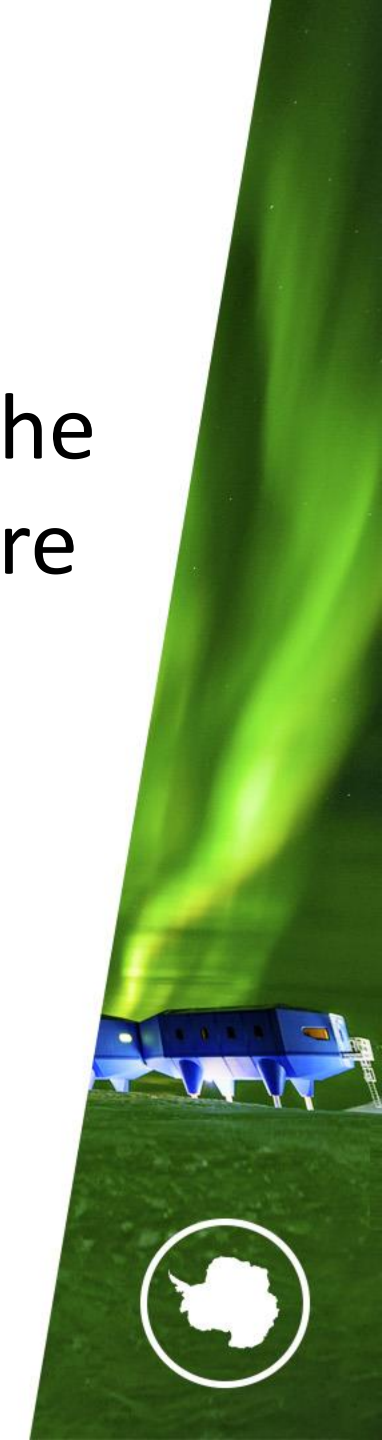
British Antarctic Survey
NATURAL ENVIRONMENT RESEARCH COUNCIL

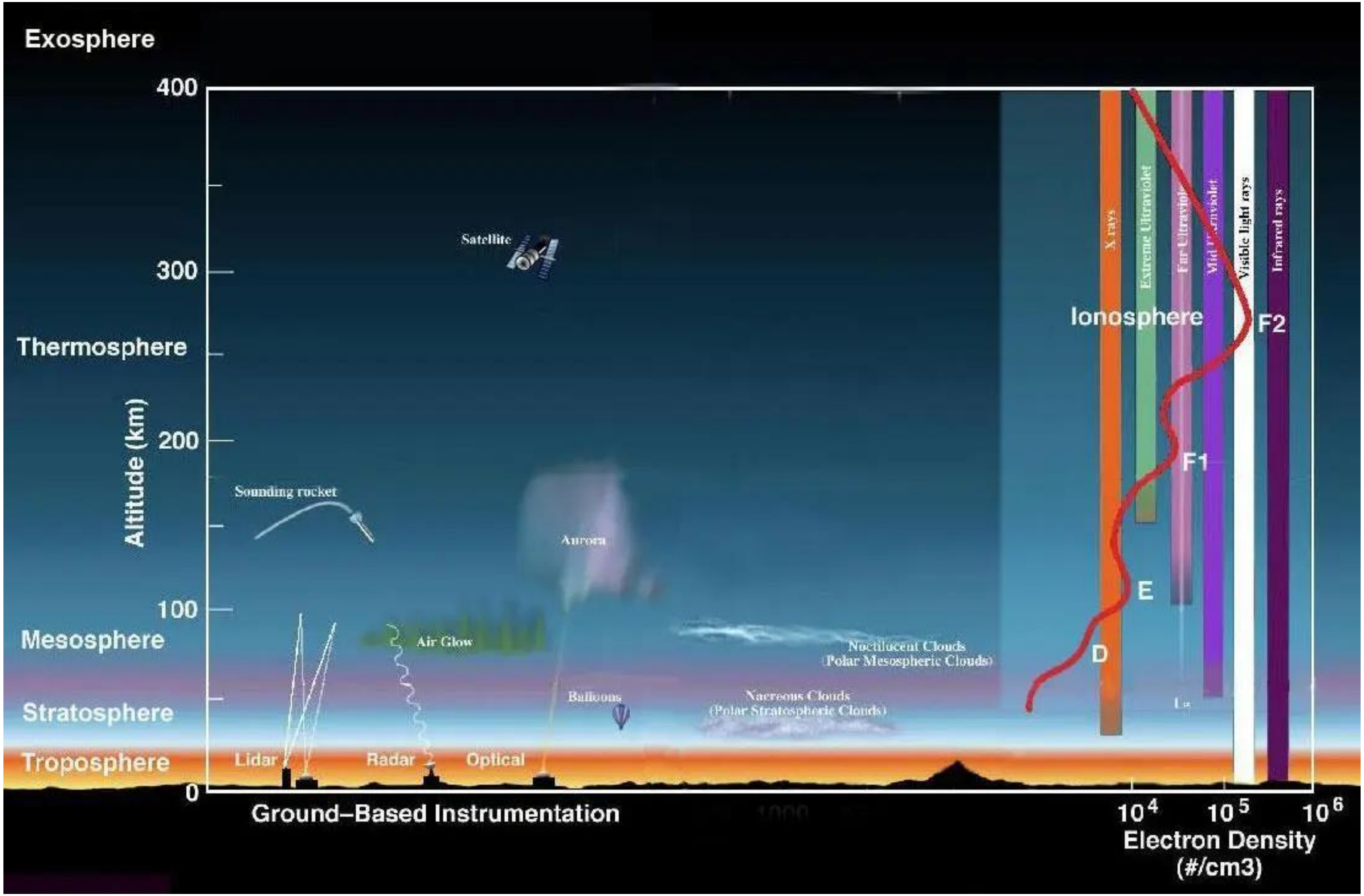
POLAR SCIENCE
FOR A SUSTAINABLE PLANET

- What is the Ionosphere?
- Why is the Ionosphere Important?
- History of the discovery of the Ionosphere
- What causes the ionosphere?
- Features and Regions of the ionosphere
- Processes in the ionosphere
- The High Latitude Ionosphere

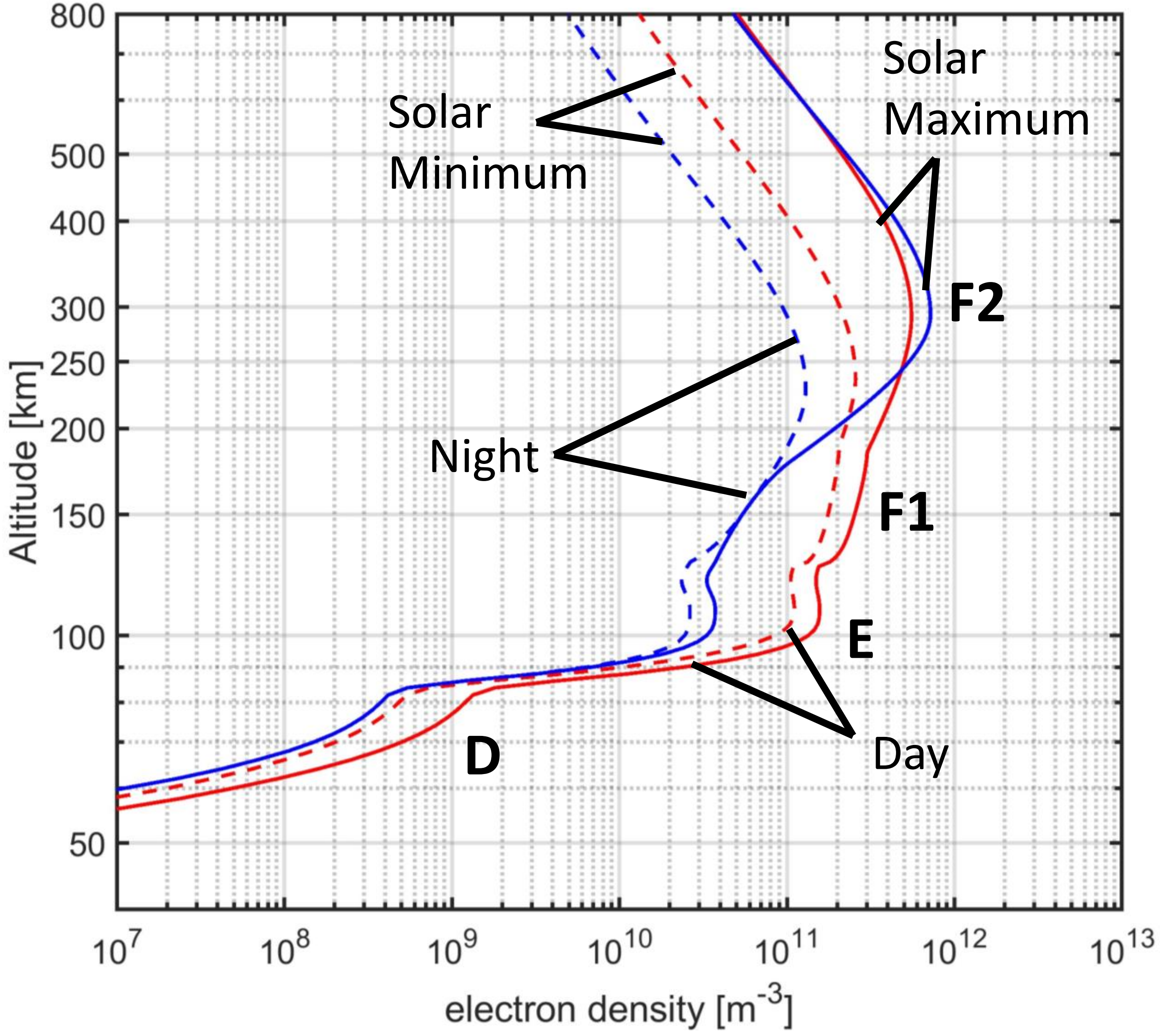


Layers of the Atmosphere




Exosphere
Thermosphere
Mesosphere
Stratosphere
Troposphere
Altitude (km)
400
300
200
100
0
Satellite
Sounding rocket
Aurora
Air Glow
Noctilucent Clouds
(Polar Mesospheric Clouds)
Balloons
Nacreous Clouds
(Polar Stratospheric Clouds)
Lidar
Radar
Optical
Ground–Based Instrumentation
X rays
Extreme Ultraviolet
Visible light rays
Infrared rays
Ionosphere
F2
F1
E
D
10 4
10 5
10 6
Electron Density
(#/cm3)

What is the Ionosphere?

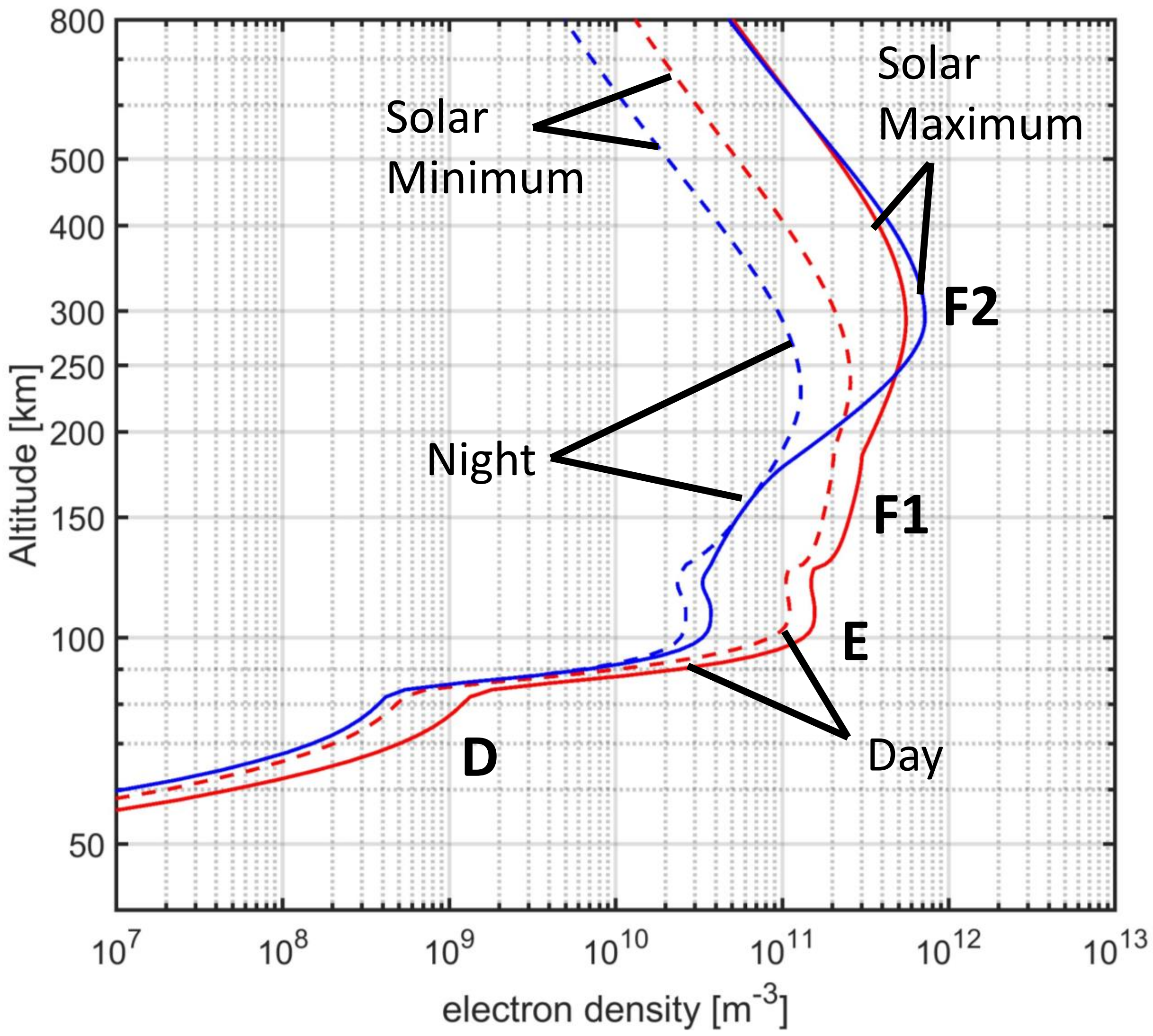


A weakly ionized plasma co-existing with the neutral atmosphere (the mesosphere and thermosphere).

Very strong solar cycle dependence
Strong diurnal variation

- F layer is two layers: F1 and F2
  - F1 layer disappears at night but F2 remains
  - Collisionless
- E layer persists at night but weaker
  - Weakly collisional
  - Where the electrical currents flow
- D layer much stronger in the daytime
  - Strongly collisional
  - Interesting ion chemistry & negative ions

What is the Ionosphere?



Composed of different layers and different regions with different processes and mechanisms dominating

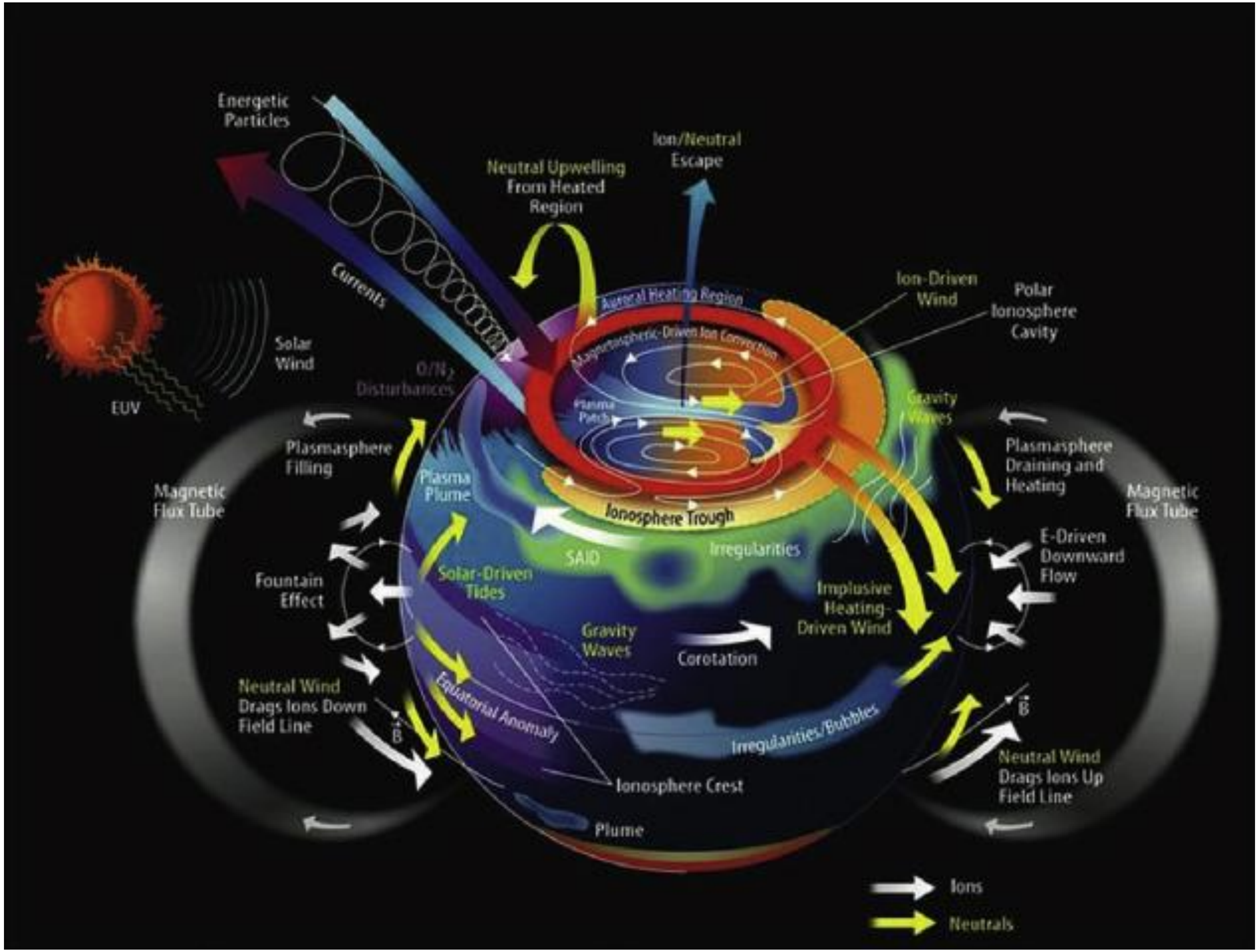
Collisionless to collisional plasma physics
Multi-scale processes

Processes include:

- Dynamics
- Kinetics
- Chemistry
- Electrodynamics

It is not a simple thing and it is impossible to cover all aspects of the Ionosphere in one hour.
(not to mention the ionospheres of other planets…)

What is the Ionosphere?



One part of the Geospace Environment

Strongly Coupled to the magnetosphere and the neutral atmosphere.

Highly important region BUT also made up of many different processes and sub-regions

Why is the Ionosphere Important?

Sometimes not seen as being ‘as exciting’ as the magnetosphere or the Sun.

BUT

Fundamentally a coupled part of the solar-terrestrial system

Gateway for observations of larger Geospace
- Understanding current systems
- Understanding of reconnection
- Understanding of substorm processes
- Energy deposition and transport via particles and waves
- Long term monitoring of geomagnetic activity
- Source of ions in the magnetosphere

Where many space weather impacts manifest – the heart of solar terrestrial physics
- GNSS signal fade/degradation
- Geomagnetically Induced Current (GIC) generation
- Atmospheric heating and satellite/space debris drag

A note on coordinate systems…

- The ionosphere coexists with a (much larger) background of neutral atmosphere.
- It is primarily generated by the Sun and responds to waves and tides that originate in the neutral atmosphere.

Therefore it makes absolute sense to work in a geographical coordinate system!

A note on coordinate systems…

- The ionosphere coexists with a (much larger) background of neutral atmosphere.
- It is primarily generated by the Sun and responds to waves and tides that originate in the neutral atmosphere.

Therefore it makes absolute sense to work in a geographical coordinate system!

- The ionosphere is a magnetised plasma, ordered by the Earth's magnetic field.
- A lot of the motion is determined by the interaction of magnetic and electric fields.
- In some regions it is strongly driven by magnetospheric processes

Therefore it makes absolute sense to work in a geomagnetic coordinate system!

A note on coordinate systems…

- The ionosphere coexists with a (much larger) background of neutral atmosphere.
- It is primarily generated by the Sun and responds to waves and tides that originate in the neutral atmosphere.

Therefore it makes absolute sense to work in a geographical coordinate system!

- The ionosphere is a magnetised plasma, ordered by the Earth's magnetic field.
- A lot of the motion is determined by the interaction of magnetic and electric fields.
- In some regions it is strongly driven by magnetospheric processes

Therefore it makes absolute sense to work in a geomagnetic coordinate system!

Question:

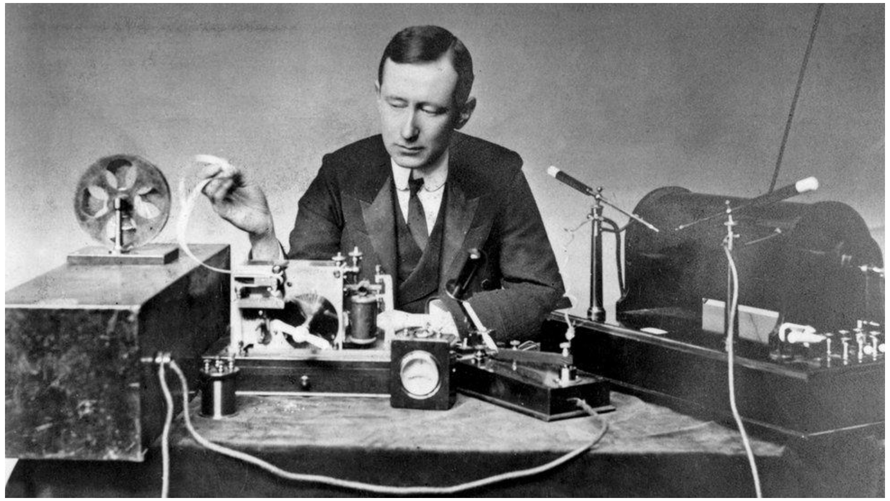
Which Coordinate system should we use???

History of the discovery of the Ionosphere

Idea of electrified layers in the atmosphere appeared in the 19th century:

- 1882 Balfour Stewart – diurnal variations in the magnetic field might be caused by electrical currents
- 1889 Arthur Schuster - presented this theory in letters to the Royal Society

In December 1901 Guglielmo Marconi conducted a radio experiment, transmitting a signal from Poldhu Cove in Cornwall, which was detected in Newfoundland Canada

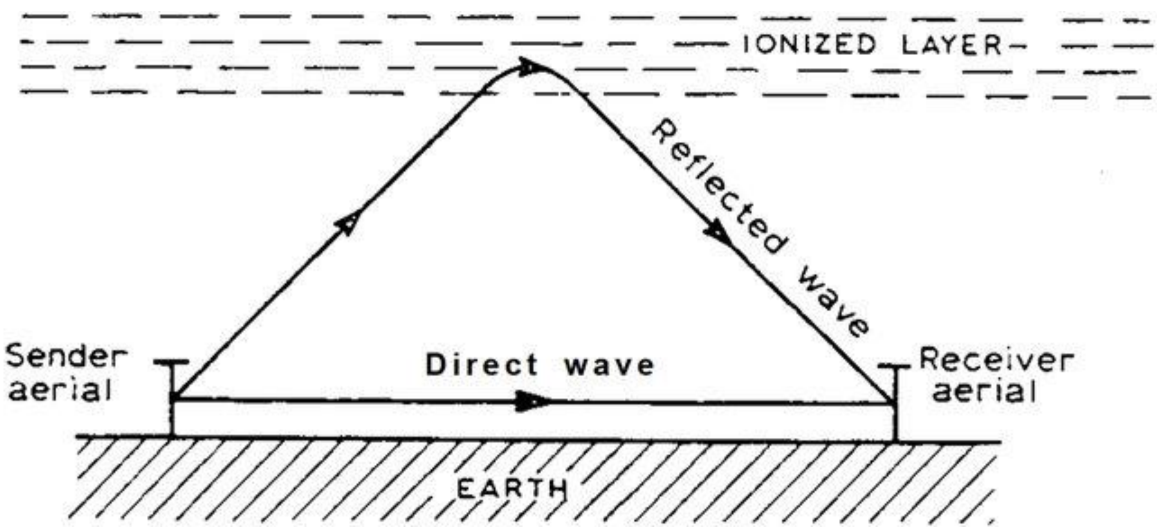


For the radio waves to be detected they must have been 'bent' as they travelled through the atmosphere. Marconi noted that signals were stronger in daylight than during night.

History of the discovery of the Ionosphere

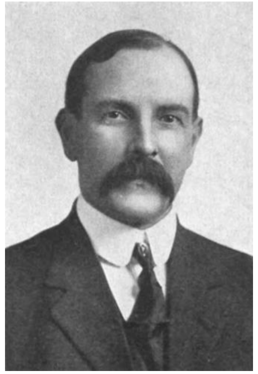
In 1902 both Arthur Kennelly and Oliver Heaviside predicted the existence of a reflective layer as an explanation for Marconi's results

The existence of the 'Heaviside layer' was shown by Edward Appleton in 1925 for which he was awarded the 1947 Nobel Prize.
Appleton bounced radio waves off the suspected layer and determined it lay at ~100 km altitude.
He later demonstrated the existence of a second layer at ~250 km.

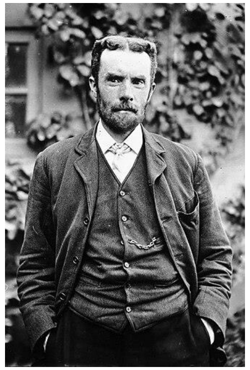


From Appleton's 1947 Nobel prize lecture

Appleton



Kennelly



Heaviside

- 'Ionosphere' was coined by Robert Watson-Watt in 1926
- adopted into common usage by ~1932
- followed the naming convention of the 'stratosphere' and 'troposphere'

Radio Wave Propagation

The discovery of the ionosphere was tied to radio propagation and many of the measurements we make of the ionosphere today come from radio or radar observations.

Radio waves passing through a charged medium will pass energy backwards and forwards between the phtons and the charge carriers.

Magnetised plasma is birefringent such that radio waves follow two separate raypaths:
O (ordinary) mode and E or X mode (extraordinary mode). There is also sometimes a Z mode.

We can define a Plasma Frequency

$$\omega_p^2 = \frac{N_e e^2}{\varepsilon_0 m_e}$$

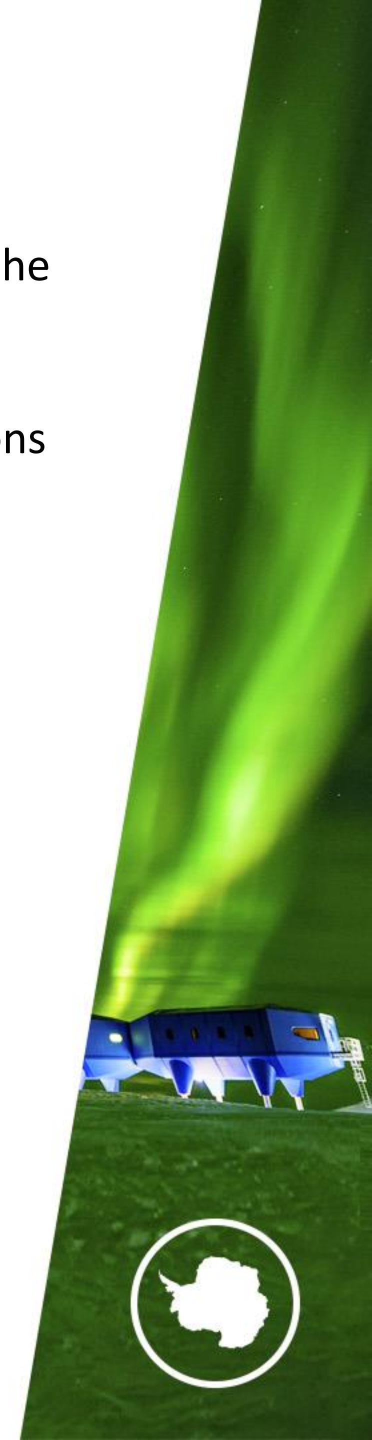
N_e	*electron concentration*
e	*charge on an election*
E	*permittivity of free space*
m_e	*Mass of an electron*

O mode waves will pass through the ionosphere if their angular frequency is higher than the plasma frequency but will reflect if their frequency is lower

$$\omega_O = \omega_p$$

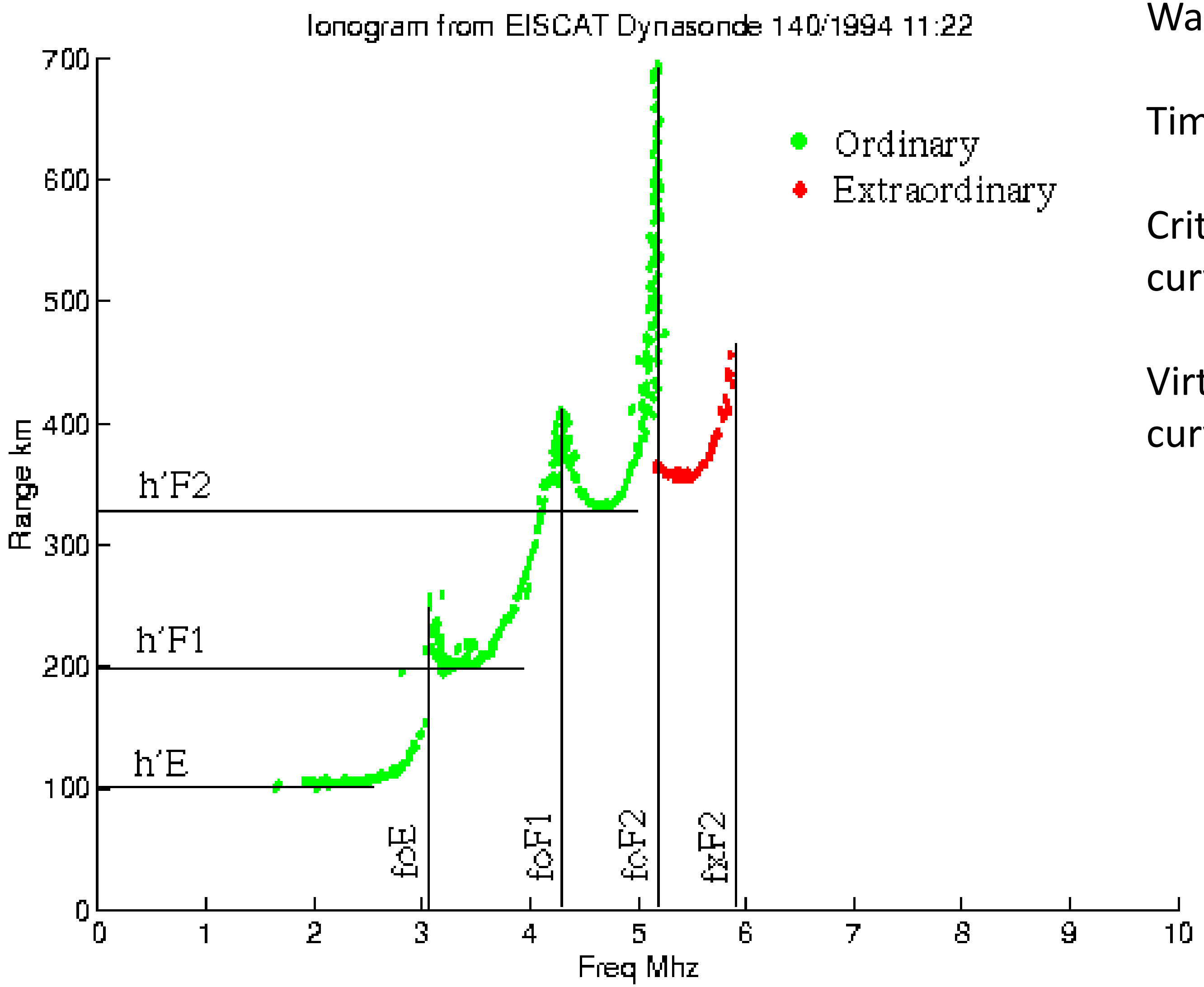
For X mode waves the critical frequency is slightly different, modified by the gyrofrequency (to account for the magnetic field). As long as the gyrofrequency is much smaller than the plasma frequency:

$$\omega_X = \omega_O + \frac{eB}{2m_e}$$



Radio Wave Propagation

A sample **ionogram** taken from an **ionosonde** (the EISCAT Dynasonde)



Waves transmitted vertically in a frequency sweep.

Time of flight gives the virtual height of the layer

Critical frequency scaled from the asymptote of the curve

Virtual height is scaled from the lowest point of the curve

Thanks to Chris Scott and Matthew Wild from the UKSSDC website

Index of Refraction $n = \frac{c}{v_p}$. in the Ionosphere

$$n^2 = 1 - \frac{X}{1 - iZ - \frac{\frac{1}{2}Y^2 \sin^2\theta}{1-X-iZ} \pm \frac{1}{1-X-iZ}\left(\frac{1}{4}Y^4 \sin^4\theta + Y^2\cos^2\theta(1-X-iZ)^2\right)^{1/2}}$$

n is the index of refraction

$$X = \frac{\omega_N^2}{\omega^2} \quad Y = \frac{\omega_H}{\omega} \quad Z = \frac{\nu}{\omega} \quad \omega_N = \left(\frac{Ne^2}{\varepsilon_0 m_e}\right)^{1/2} \quad \omega_H = \frac{e|B|}{m_e}$$

ω = the angular frequency of the radar wave,

$Y_L = Y\cos\theta$, $Y_T = Y\sin\theta$,

θ = angle between the wave vector $\bar{k}$ and $\bar{B}$,

$\bar{k}$ = wave vector of propagating radiation,

$\bar{B}$ = geomagnetic field, N = electron density

e = electronic charge, m_e = electron mass, ν = electron collision frequency

and ε_0 = permittivity constant.

Developed independently by
Edward Appleton
Douglas Hartree
Wilhelm Altar

First calculated in 1926 by Altar (while working with Appleton)

German Physicist H. K. Lassen also developed it but published only in German…

The **Altar-Appleton-Hartree Equation** describing the refractive index for electromagnetic wave propagation in a cold magnetised plasma

Radio Propagation in the Ionosphere

Index of Refraction (no **B** field)

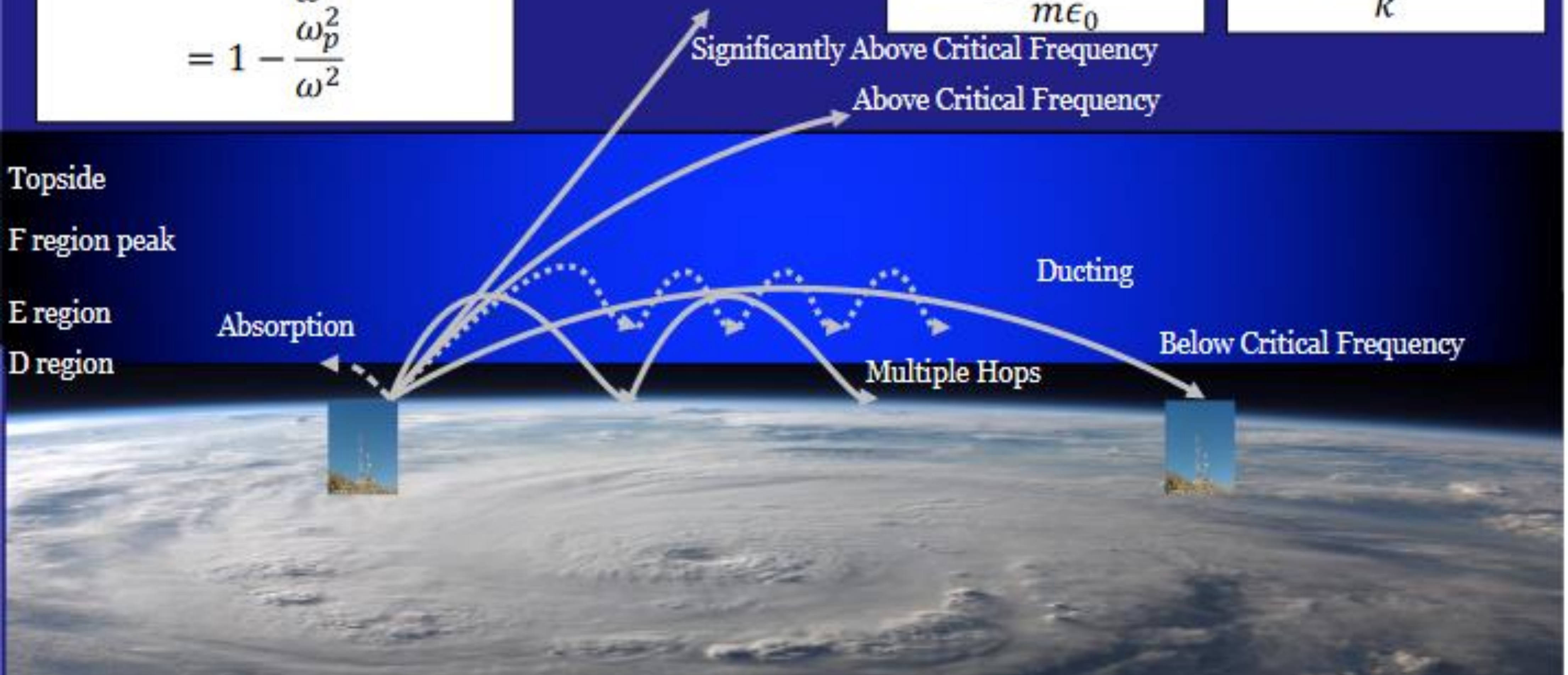
$$n^2 = \frac{c^2k^2}{\omega^2} = 1 - \frac{\omega_p^2}{\omega^2}$$

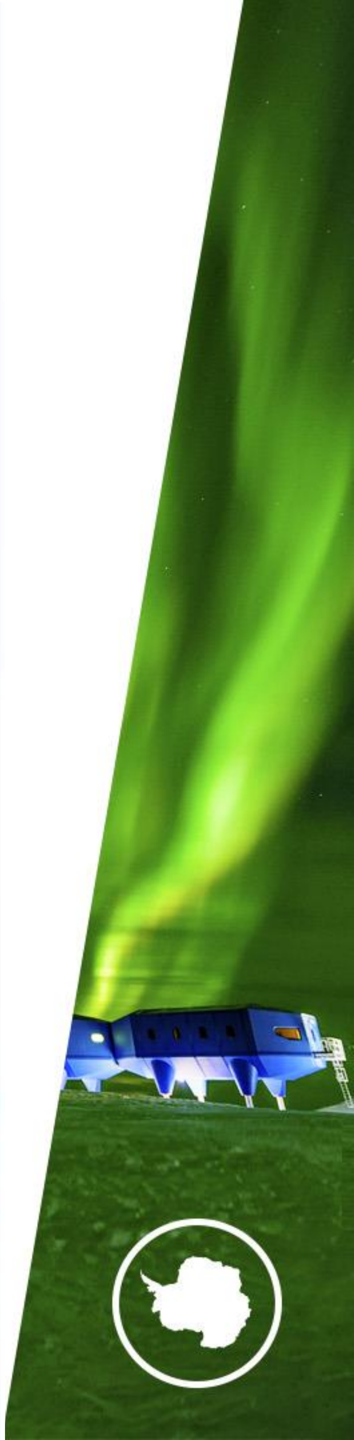
Plasma Frequency

$$\omega_p^2 = \frac{n_0 e^2}{m\epsilon_0}$$

Phase Velocity

$$V_{ph} = \frac{\omega}{k}$$





What causes the Ionosphere?

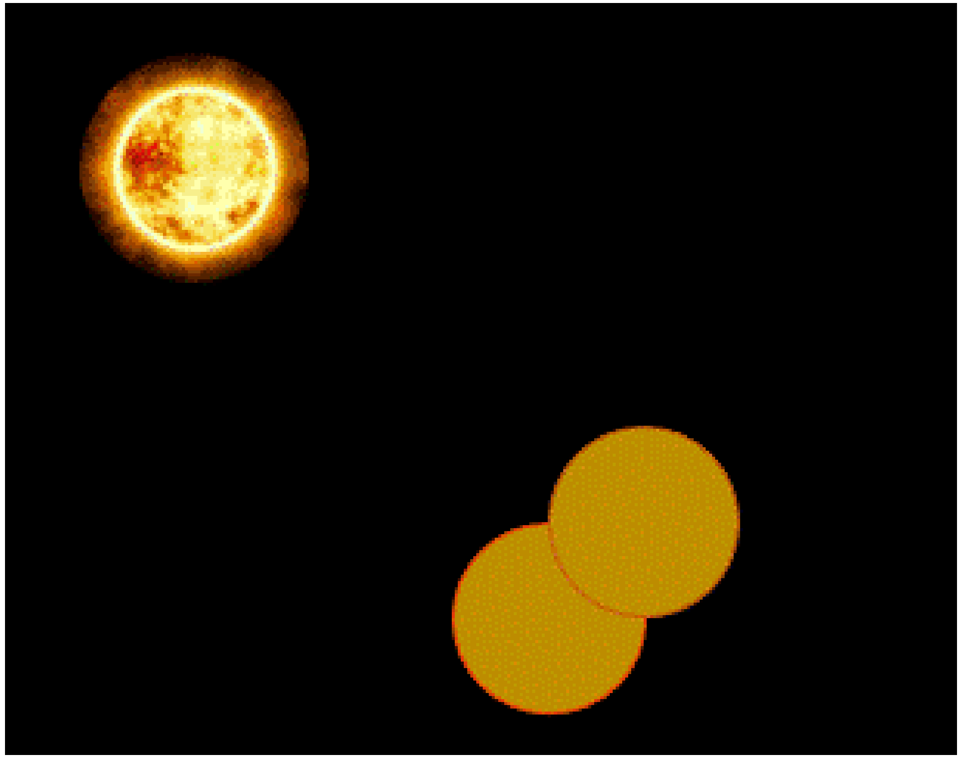
Balance between production and loss processes

Photo-ionisation
– mostly EUV and some X-ray

What other production processes are there?

Recombination
– electrons remerge with ions

What other loss mechanism are there?
What might happen during recombination?



EUV from the Sun provides enough energy to remove an electron from the molecule. High energies involved lead to electron gas having a high temperature (1000+ K)

What causes the Ionosphere?

Sidney Chapman developed a formula in 1931 to predict the form of an ionospheric layer based on 4 terms

$$q = \eta\sigma n I$$

The rate of production depends on the intensity of ionizing radiation (*I*), the concentration of the neutral species (n), the absorption cross section (σ), and the ionization efficiency (η).

The Chapman production formula relies on important assumptions:

- Single species atmosphere exponentially distributed with a constant scale height
- No variations in the horizontal plane
- Solar radiation is absorbed in proportion to the concentration of gas particles
- The absorption coefficient is constant – monochromatic radiation

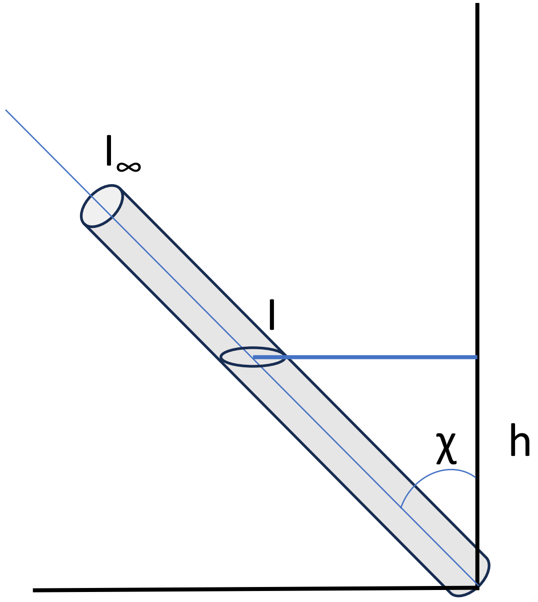
$$q(\chi, h) = q_0 \exp(1 - z - \sec(\chi)\exp(-z))$$

$$z = \frac{h - h_0}{H}$$

$$H = \frac{k_B T}{mg}$$

$$q_0 = \eta I_\infty e H$$

χ is the solar zenith angle
q_0 is the maximum production for $\chi = 0$
h_0 is the height of maximum production for $\chi = 0$
H is the scale height of the atmosphere



What causes the Ionosphere?

We can assume that electrons recombine directly with positive ions ($X^+ + e \rightarrow X$) in which case we can define a Loss term to balance the production rate.

$$L = \alpha[X^+]N = \alpha N^2$$

Where N is the electron density (equal to the ion density [X^+]) and α is the recombination coefficient

In the case of equilibrium

$$q = \alpha N^2$$

Alpha – Chapman Layer

$$q = \beta N$$

Beta – Chapman Layer

Electrons can also be lost via attachment to form negative ions.

There is also the case where positive ions are balanced by electrons and negative ions

$$q = \alpha_e N_e N_+ + \alpha_i N_- N_+$$

$$q = (1 + \lambda)(\alpha_e + \lambda\alpha_i)N_e^2$$

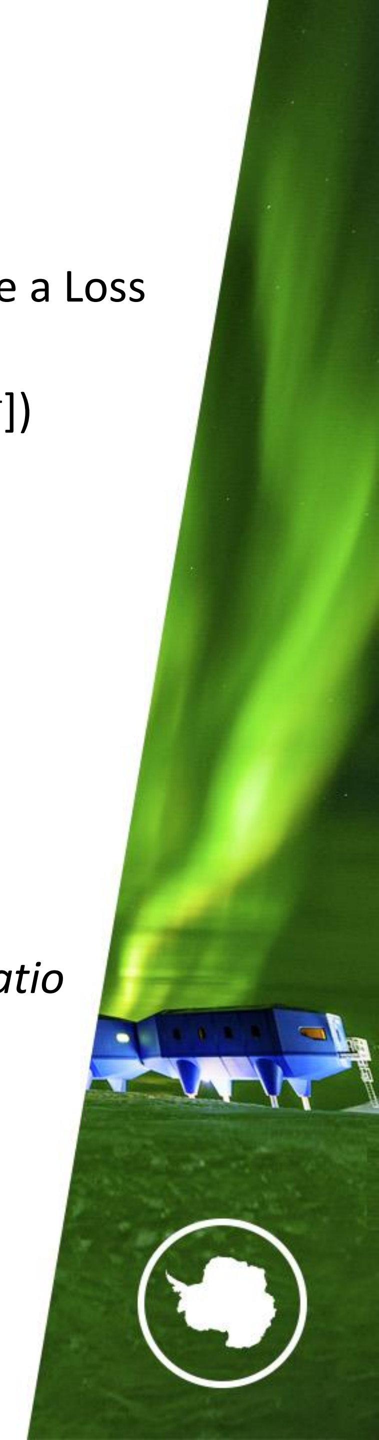
$$\lambda = \frac{N_-}{N_e}$$

Negative ion ratio

In general:

$$\frac{dN_e}{dt} = q - \alpha N^2 - \beta N$$

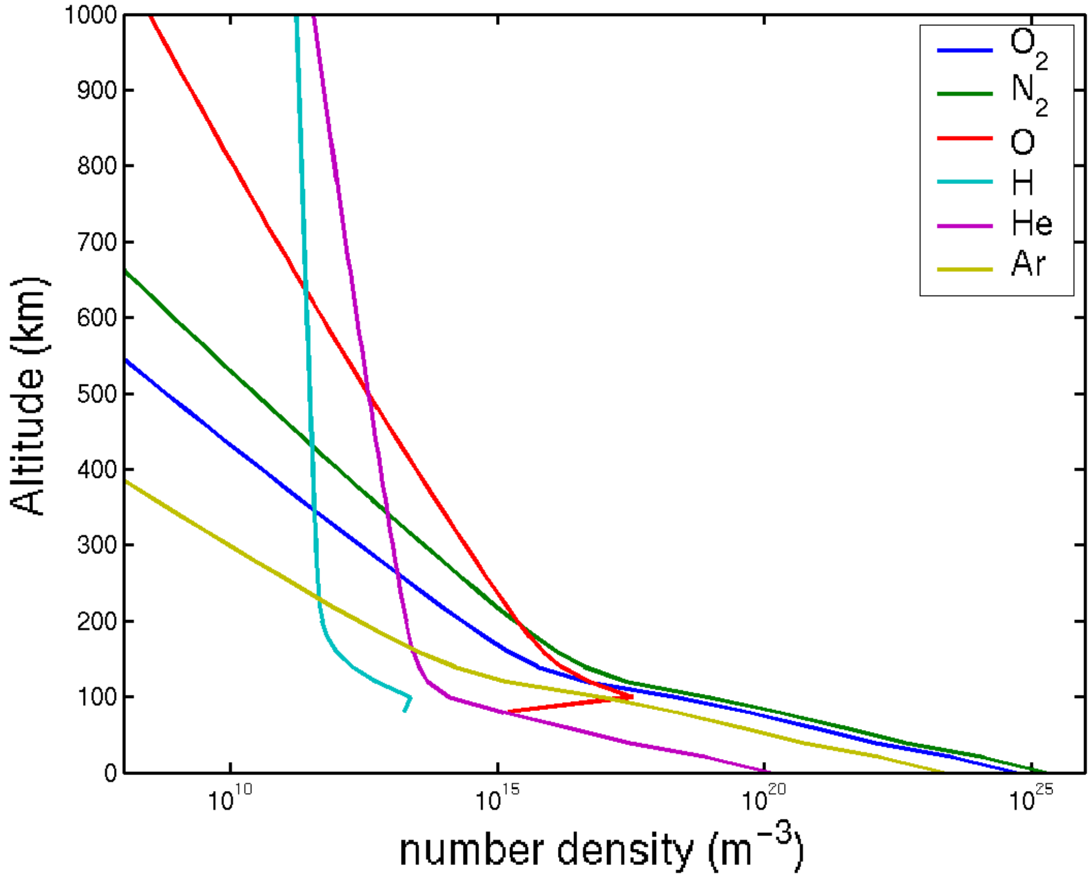
α and β can vary with altitude
This ignores other local loss and production via transport terms.



What causes the Ionosphere?

Composition!

The ionosphere comes from the neutral atmosphere (Thermosphere/Mesosphere) and so its structure depends on the atmospheric composition.



Below 90 km the atmosphere is well mixed.

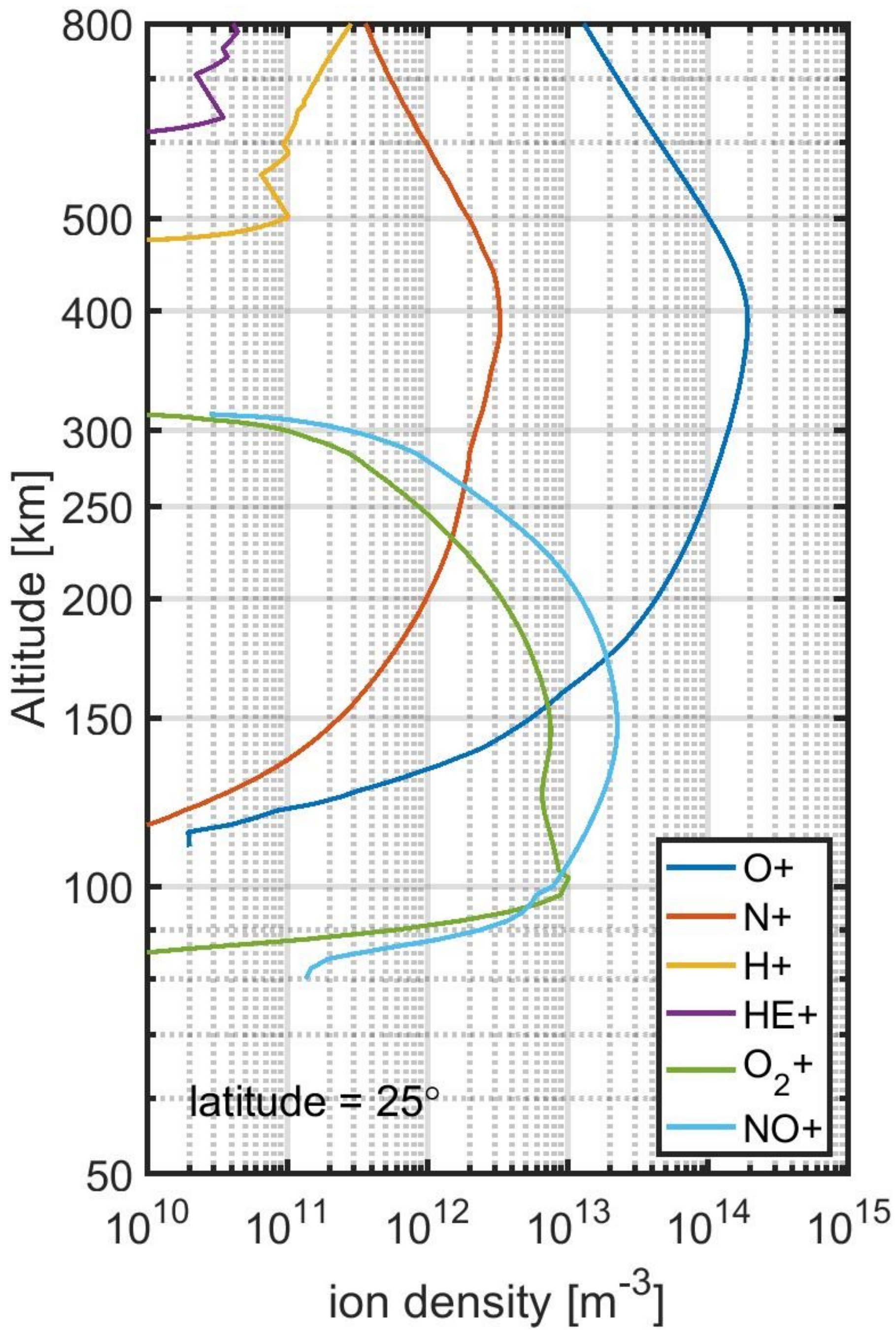
Above that height (the turbopause) atmospheric constituents are sorted due to gravity.

Neutral Composition

Ionization energy from solar radiation spectrum

Ionospheric Composition and vertical structure

What causes the Ionosphere?



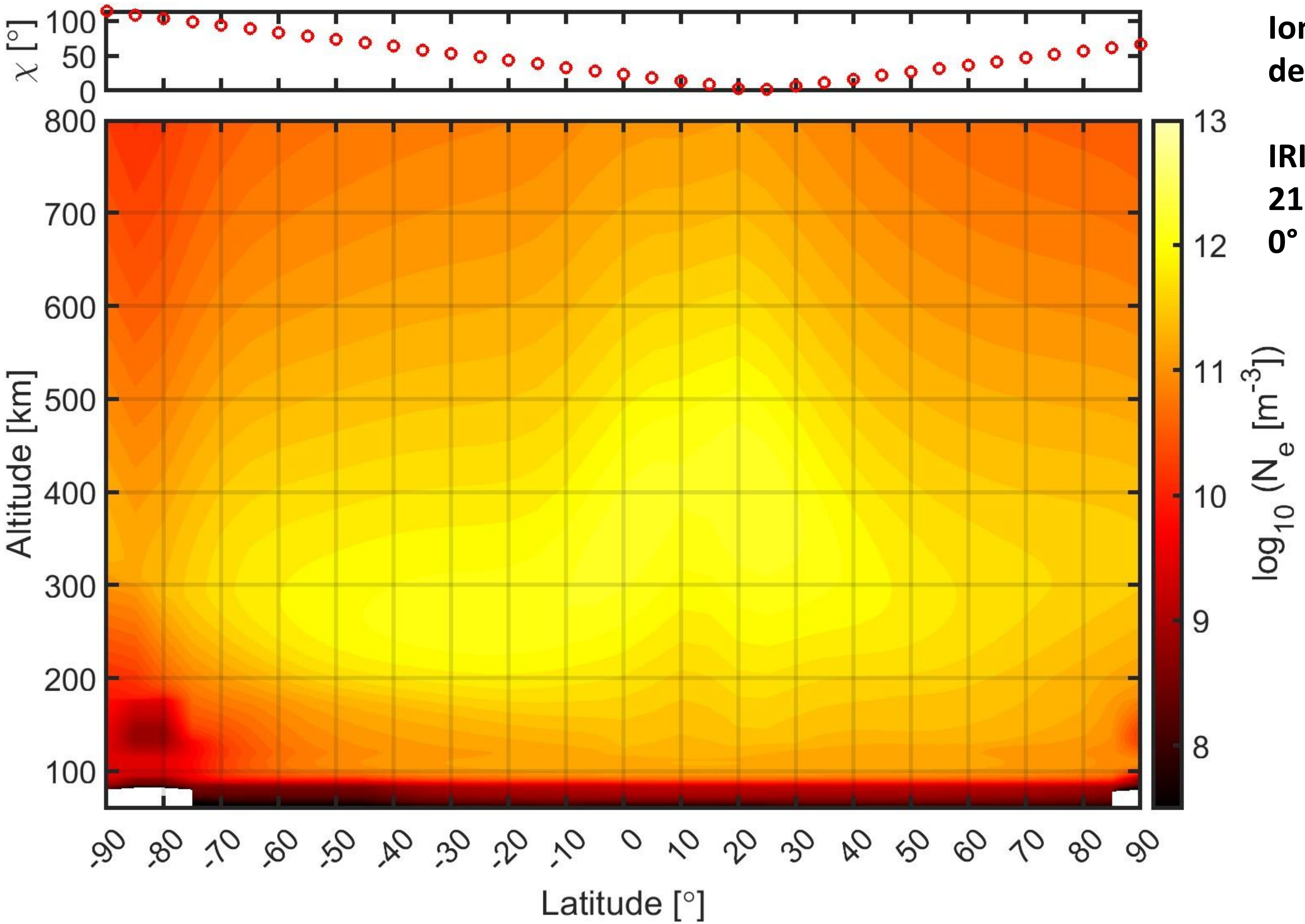
An example of ion composition at low latitudes, taken under the sub-solar point (maximum illumination).

From IRI-2020 (International Reference Ionosphere)

21 June 2001 (solstice near solar max)
Base altitude resolution = 2km, smoothed to 10 km above 150 km

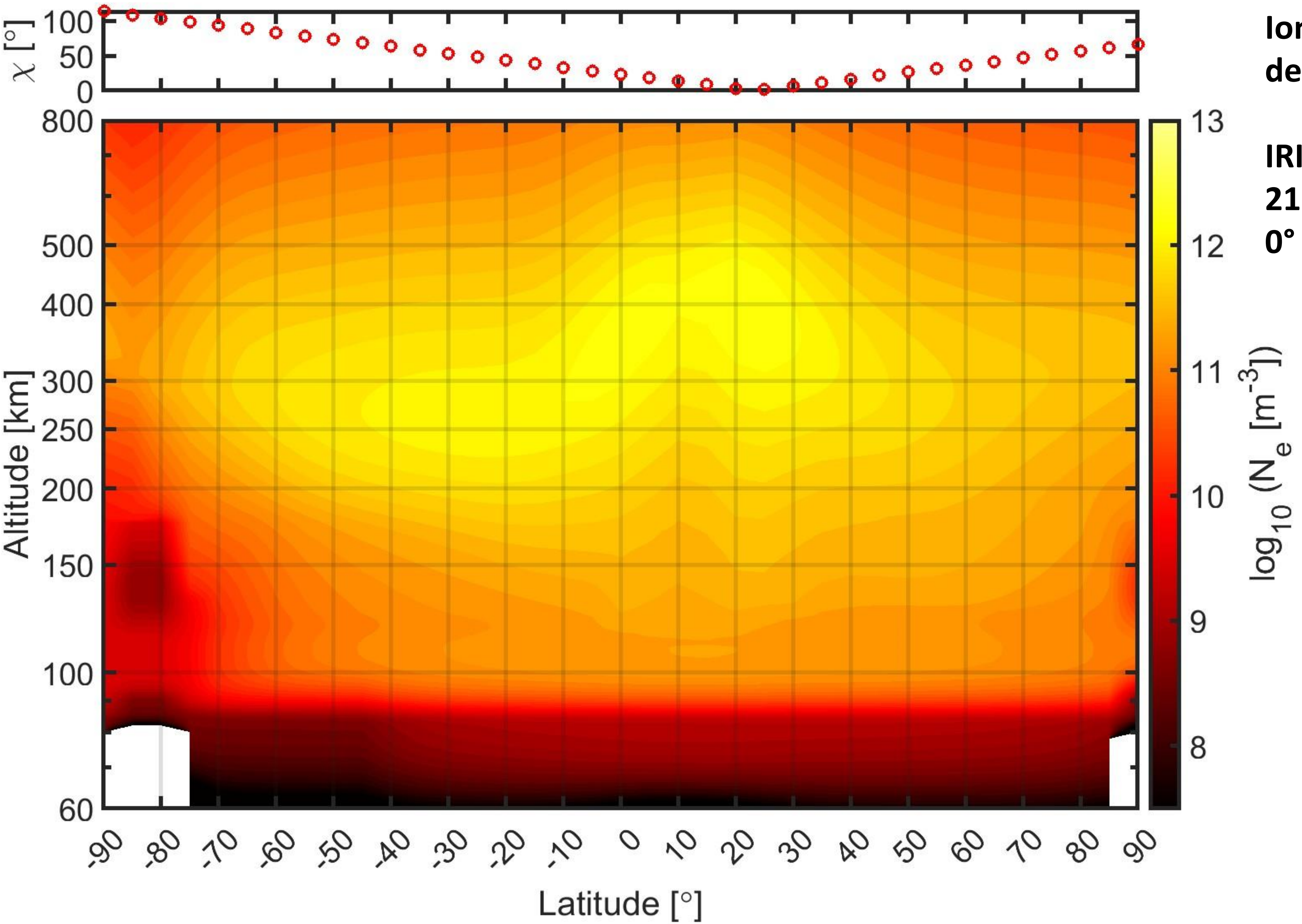
Input proxies for solar activity were:
SSN = 110.2
F10.7 = 206.9

These are the dominant species but others also exist at much lower concentrations.
Complex cluster ions can form and significant negative ions in the lower ionosphere



Ionospheric electron density

IRI-2020
21 June 2001
0° longitude

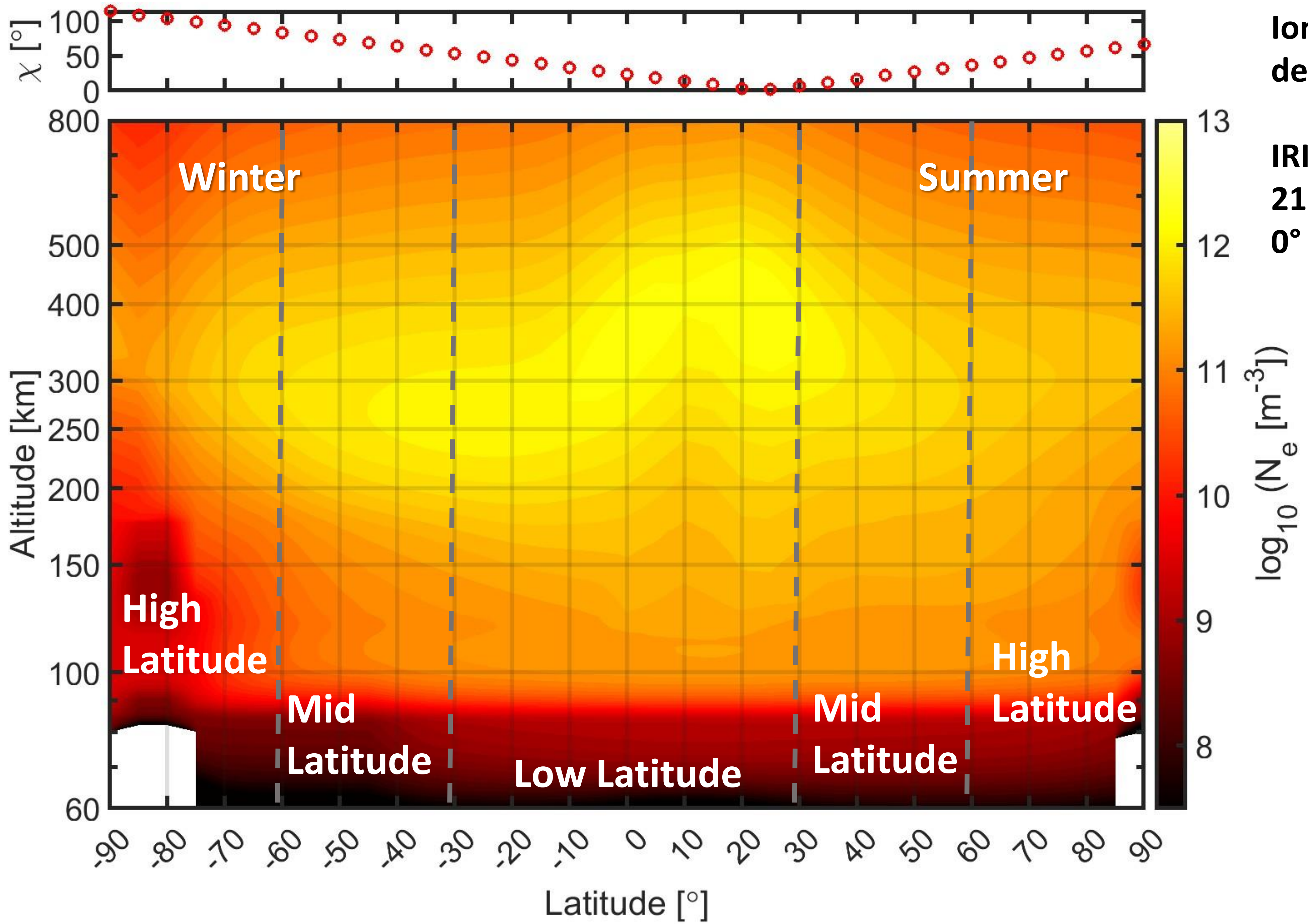


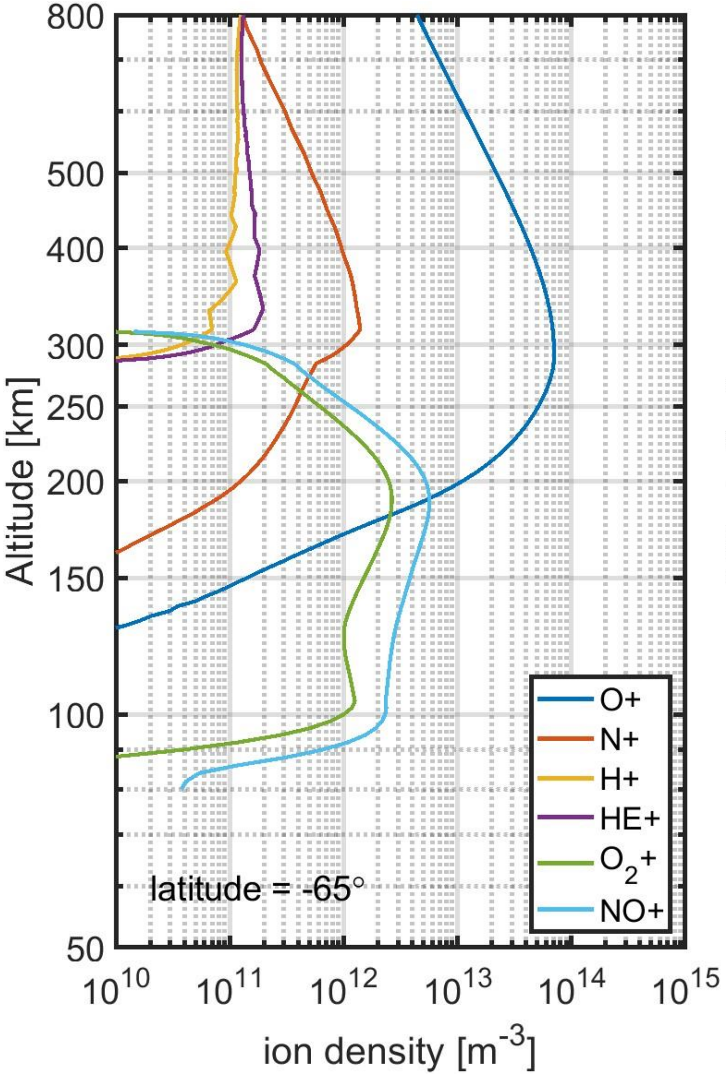
Ionospheric electron density

IRI-2020
21 June 2001
0° longitude

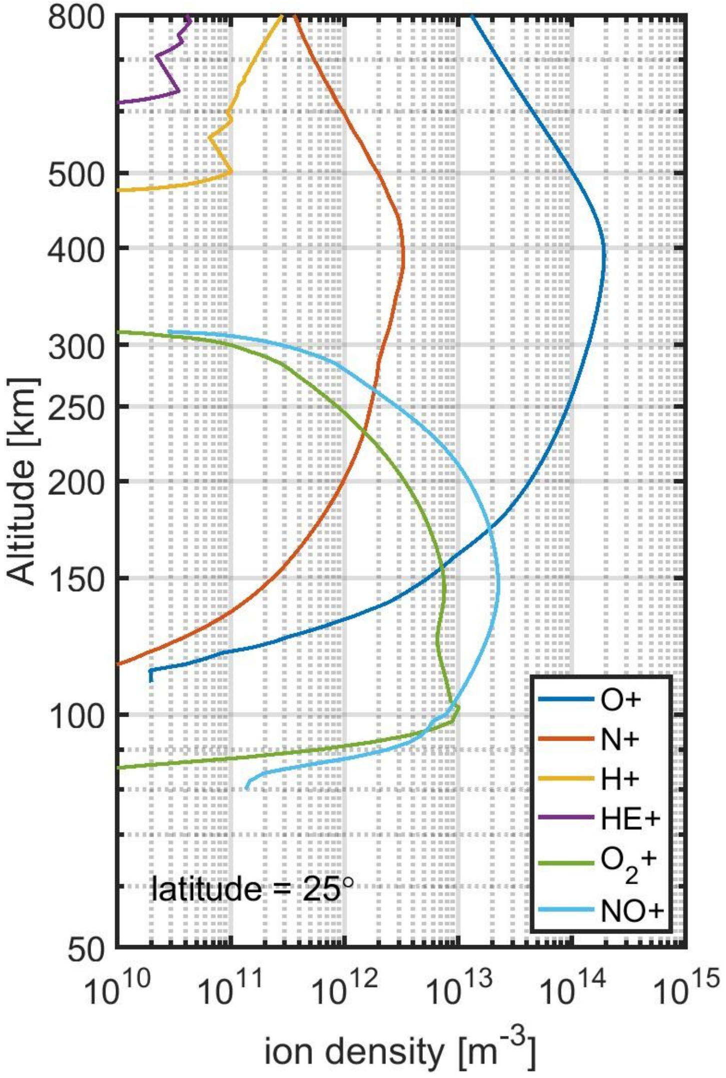
Ionospheric electron density

IRI-2020
21 June 2001
0° longitude

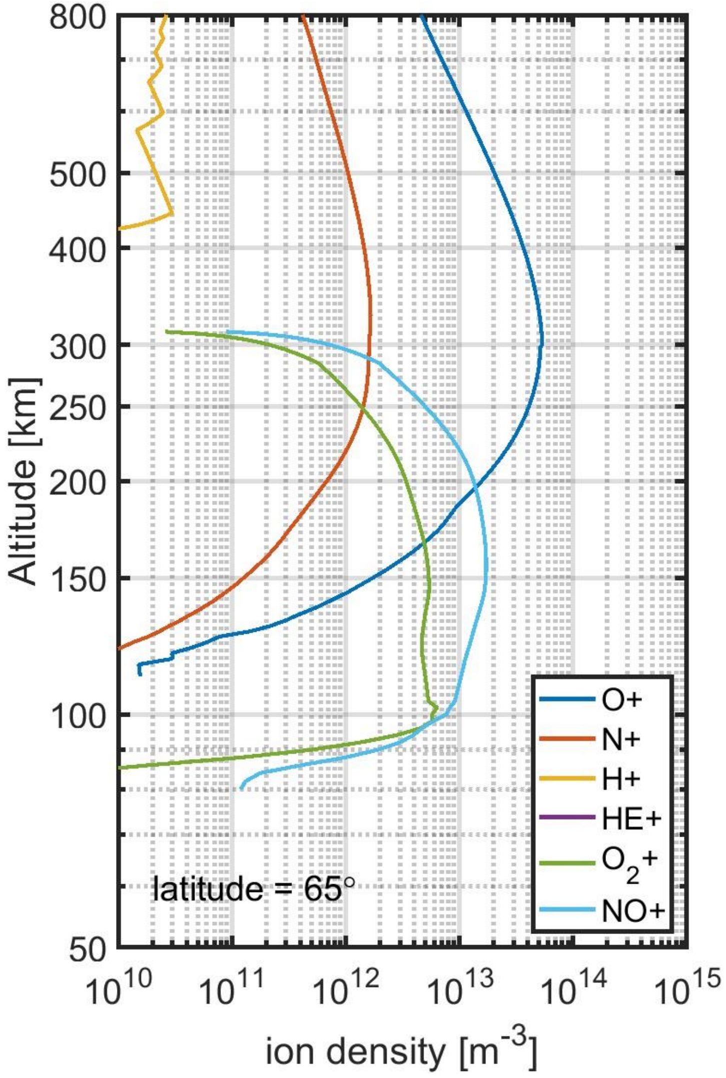




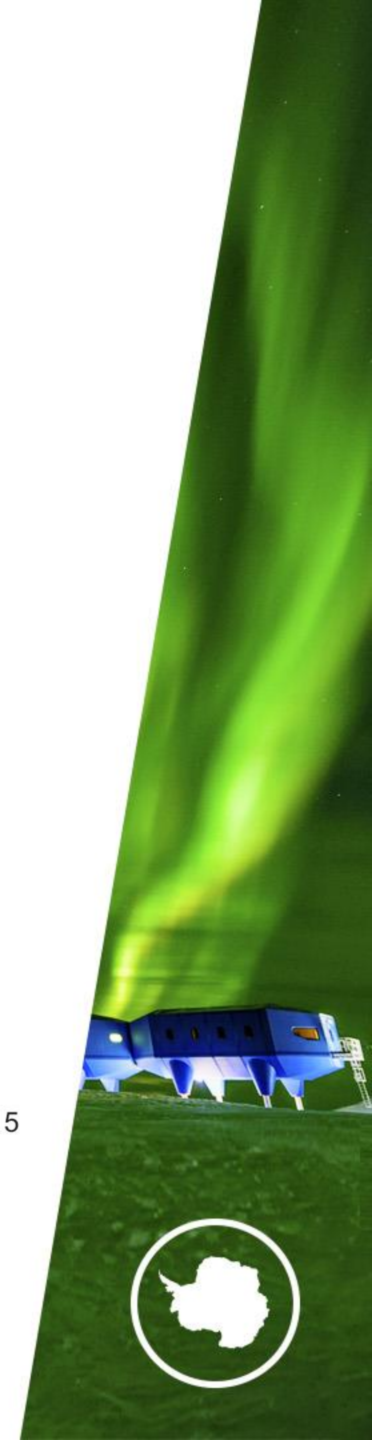
High latitude night



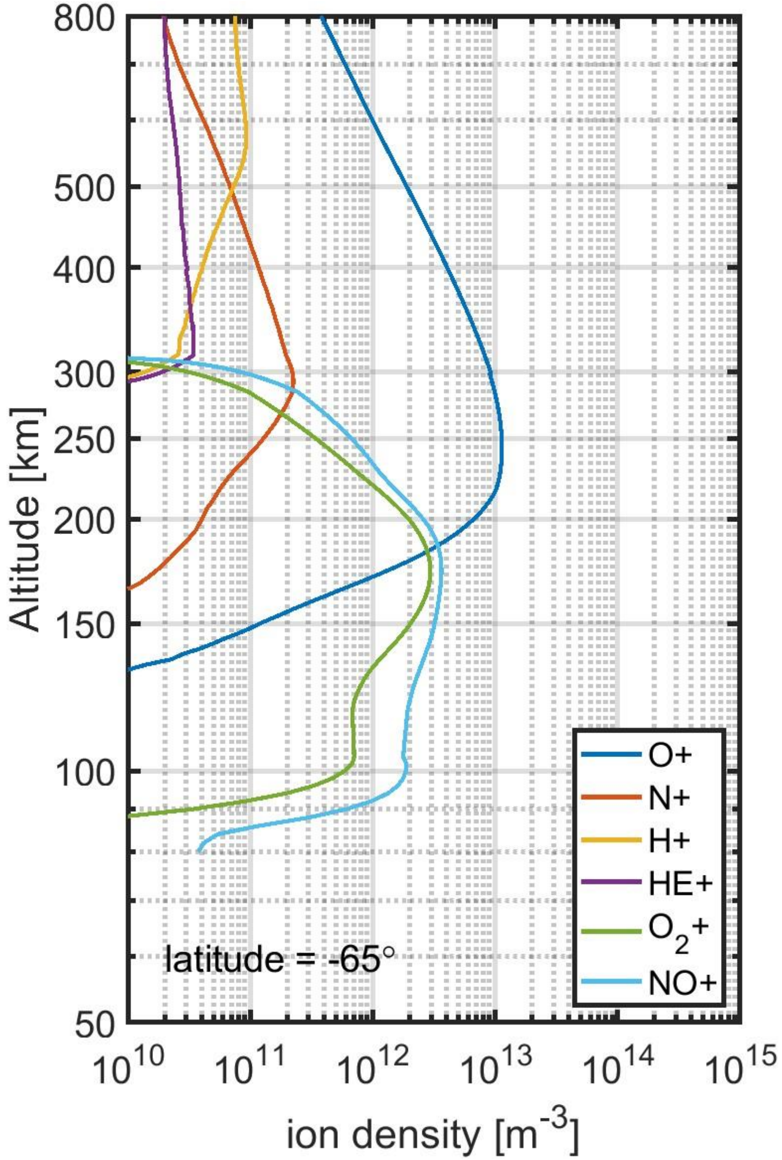
Low latitude day



High latitude day

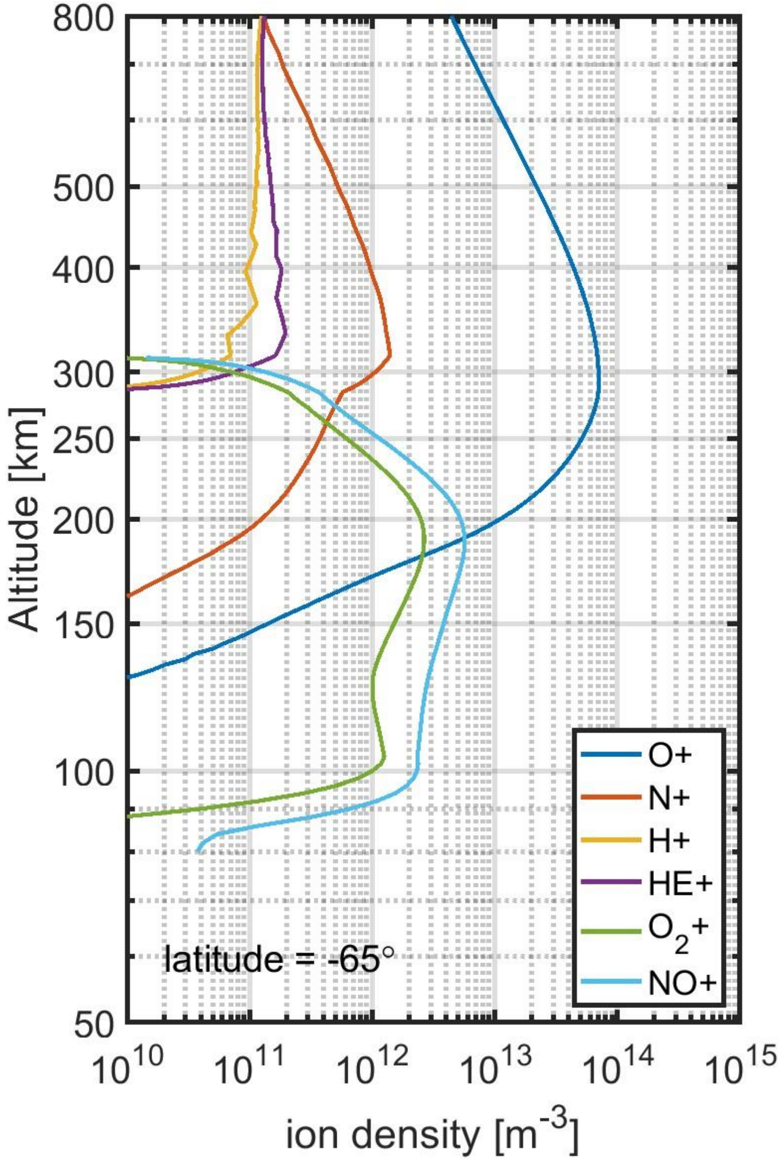


Solar Minimum - night



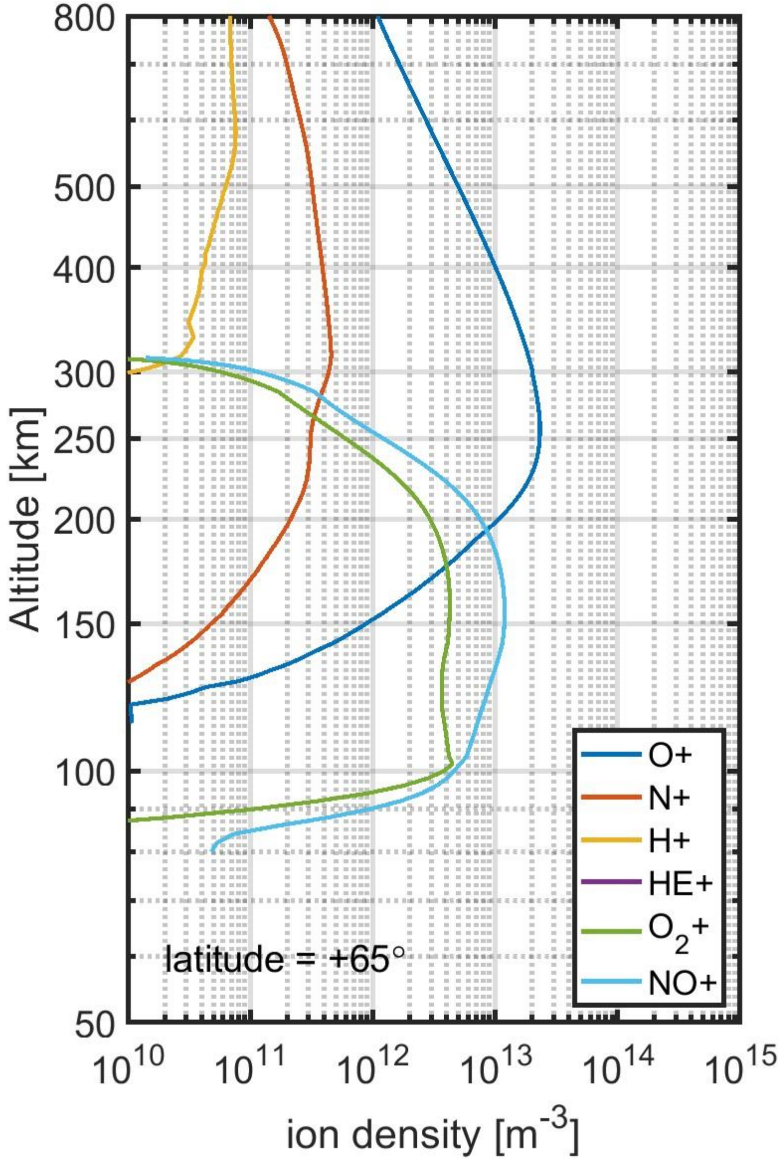
21 June 2008
SSN = 3.2
F10.7 = 67

Solar Maximum - night



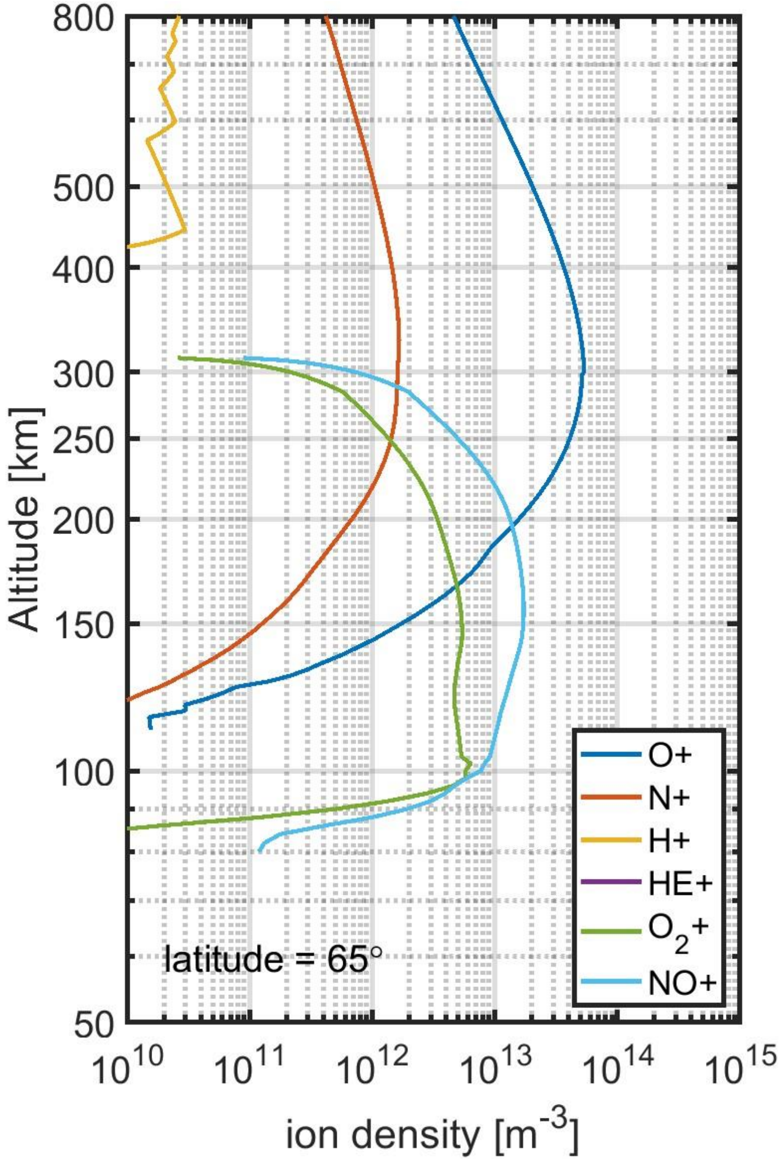
21 June 2001
SSN = 110.2
F10.7 = 206.9

Solar Minimum - day

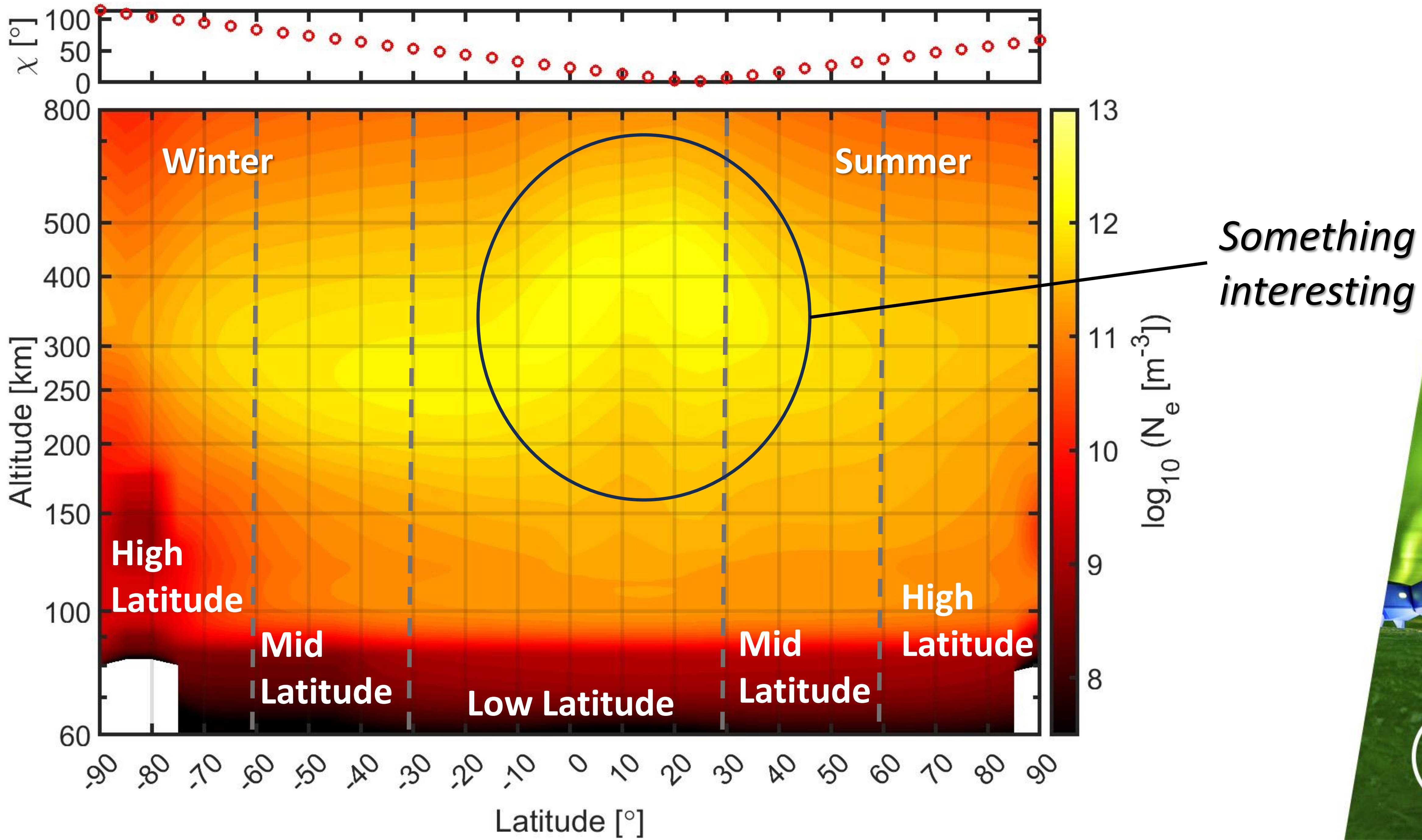


21 June 2008
SSN = 3.2
F10.7 = 67

Solar Maximum - day



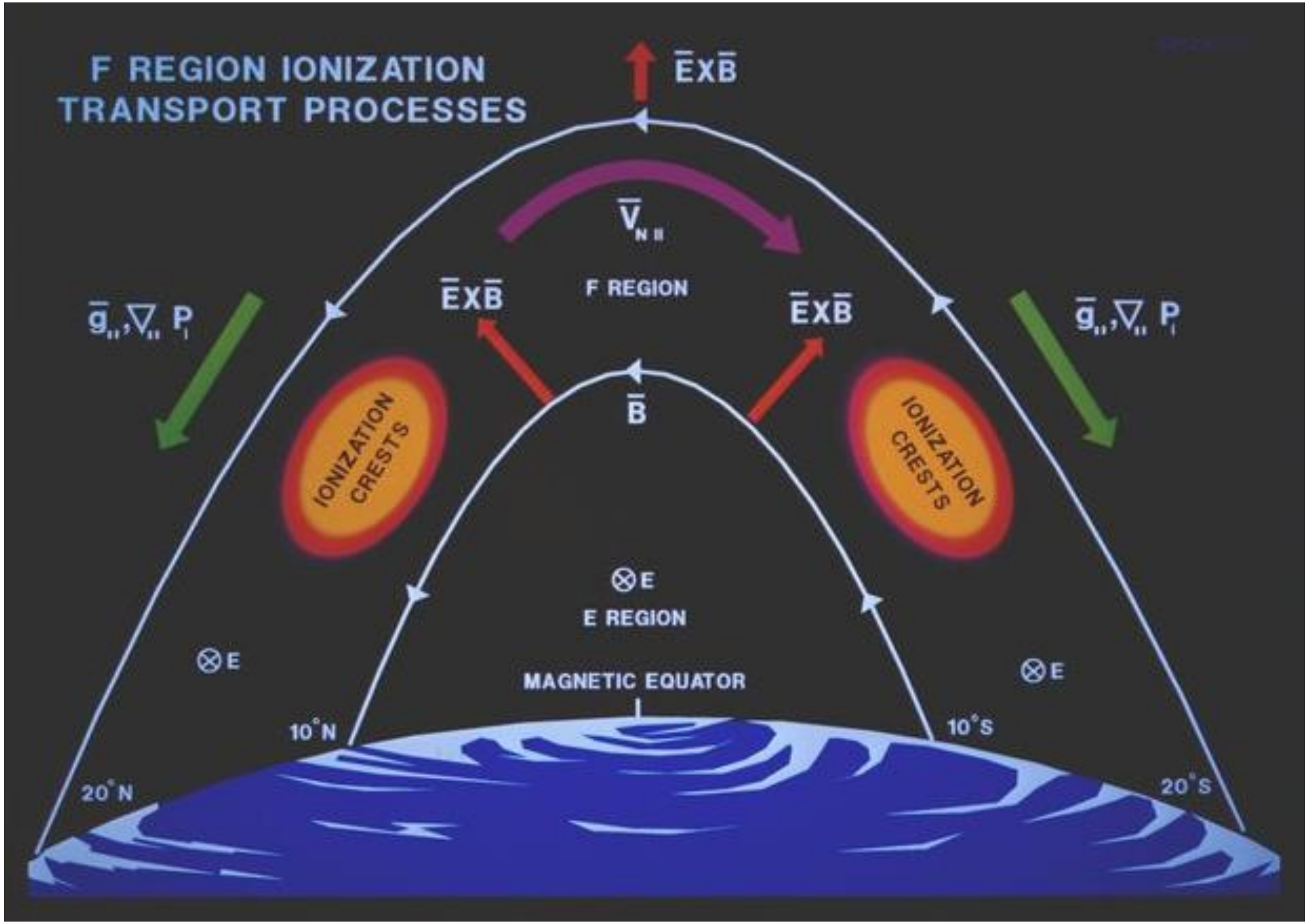
21 June 2001
SSN = 110.2
F10.7 = 206.9


χ [°]
100
50
0
800
500
400
300
250
200
150
100
60
Altitude [km]
Winter
Summer
High
Latitude
Mid
Latitude
Low Latitude
Mid
Latitude
High
Latitude
-90
-80
-70
-60
-50
-40
-30
-20
-10
0
10
20
30
40
50
60
70
80
90
Latitude [°]
13
12
11
10
9
8
log10 (Ne [m-3])
Something
interesting

Question:

How/Why does plasma flow in the ionosphere?

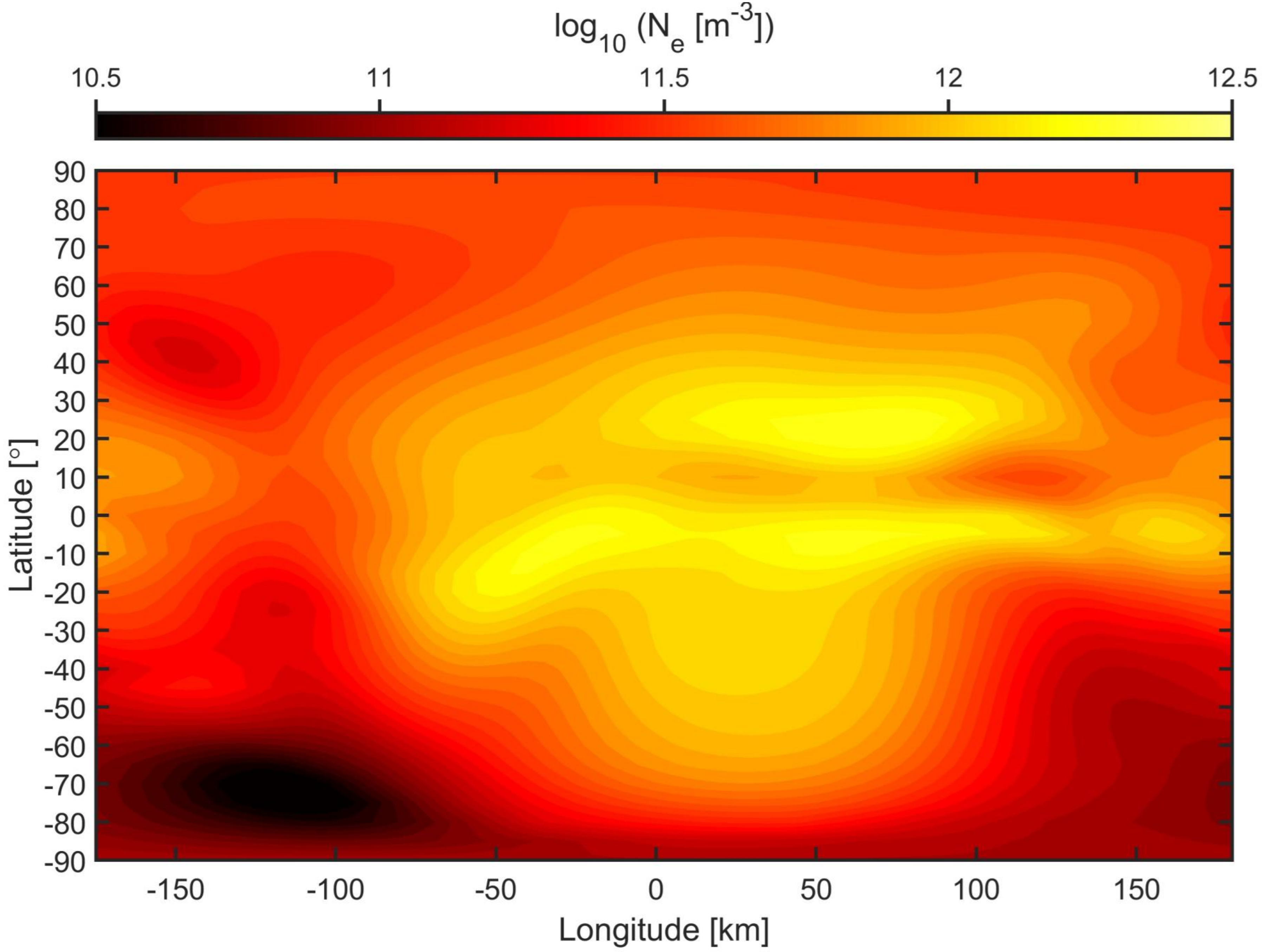
The Equatorial Anomaly – electrodynamics!



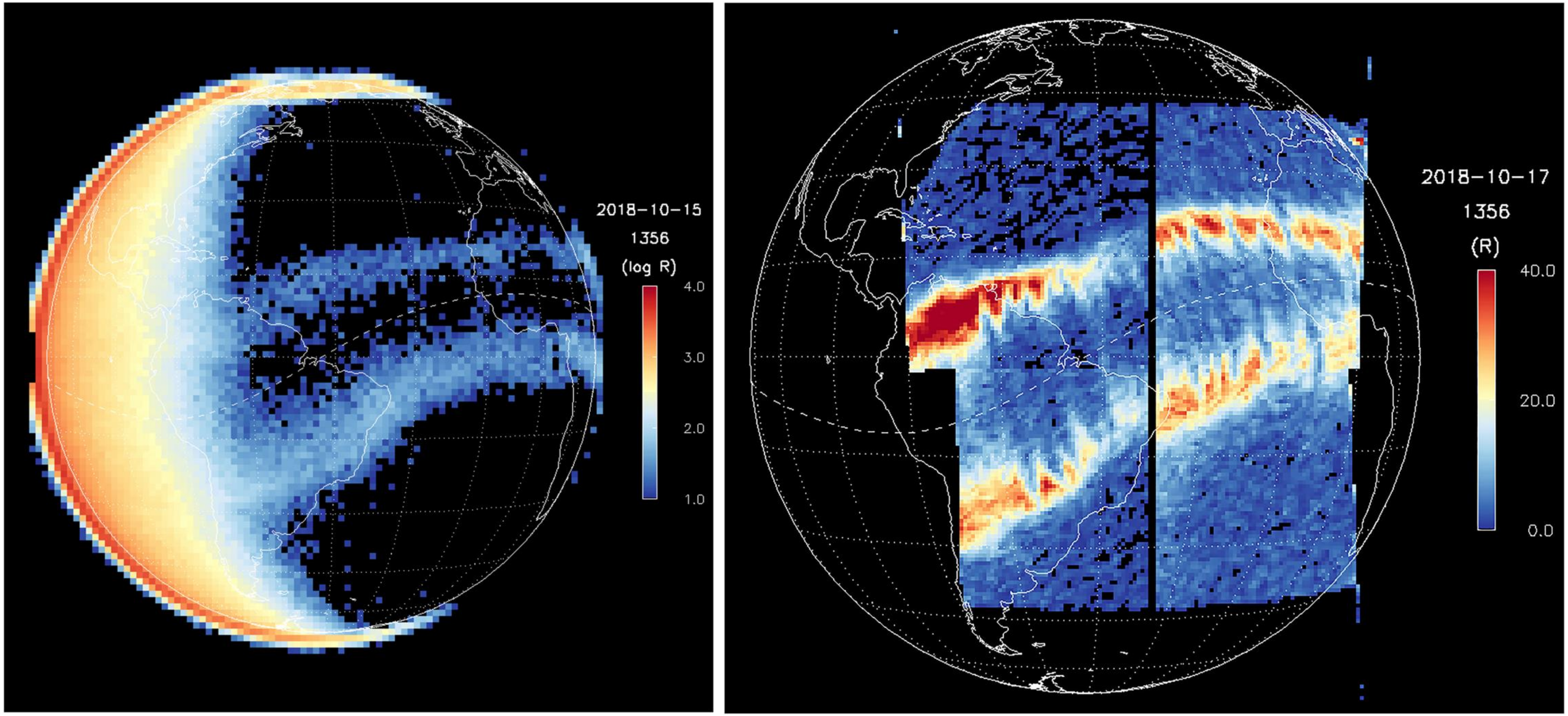
"The plasma moves upward due to E x B drift and then diffuses along the magnetic field to form two crests with maximum ionization density near ±15° magnetic latitude and minimum ionization at the magnetic equator. "

- Convection raises plasma.
- Plasma falls under gravity, bound to the field
- Depletion over the equator

*Anderson, D. N., Decker, D. T., and Valladares, C. E.: Global **theoretical** ionospheric model (GTIM) in Solar-Terrestrial Energy Program: Handbook of Ionospheric Models, National. Oceanic and Atmos. Admin, Boulder, CO, 133–152, 1996 (21)*



Can see it in the IRI model spread in longitude across the dayside



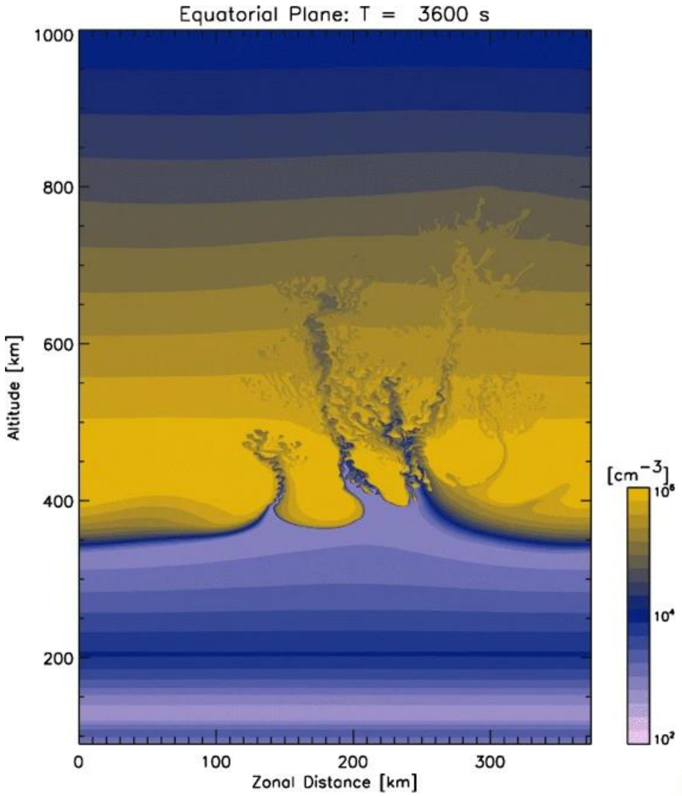
Even better we can see it in real data from the Global-scale Observations of the Limb and Disk (GOLD) imager

135.6-nm emission of atomic oxygen on 15 October 2018 and 17 October 2017
Also known as the Appleton Anomaly (our old friend!)

Eastes, R. W., Solomon, S. C., Daniell, R. E., Anderson, D. N., Burns, A. G., England, S. L., et al. (2019). Global-scale observations of the equatorial ionization anomaly. Geophysical Research Letters, 46, 9318–9326. https://doi.org/10.1029/2019GL084199

Yokoyama, T. A review on the numerical simulation of equatorial plasma bubbles toward scintillation evaluation and forecasting. Prog Earth Planet Sci **4**, 37 (2017). https://doi.org/10.1186/s40645-017-0153-6

Regions of the Ionosphere



Low Latitude: -30° to +30° latitude
primarily due to solar ionisation

- Instabilities can form leading to plasma bubbles on the nightside (~100 km scale).
- Ion recombination forms a low density layer at lower altitudes.
- Layer rises into the high density above
- Bubbles are turbulent and irregular
- Causes degradation of GNSS signals via changes in refractive index

Associated Phenomenon: the equatorial electrojet

Mid Latitude: 30° to 60° latitude
more variable illumination due to bigger variability in solar zenith angle
some additional ionisation from energetic charged particles from radiation belts

Overall quieter, fewer instabilities BUT some interesting composition effects due to reaction rates in the F layer
Good region to identify Travelling Ionospheric disturbances
- the ionospheric component of atmospheric gravity waves
- multiple scales from small, through medium (MSTID) to large scale (LSTID)
- generated by many sources, some local, some remote
- perturbations in the plasma density, temperature and flow, travel incredibly long distances

Sources of TID:

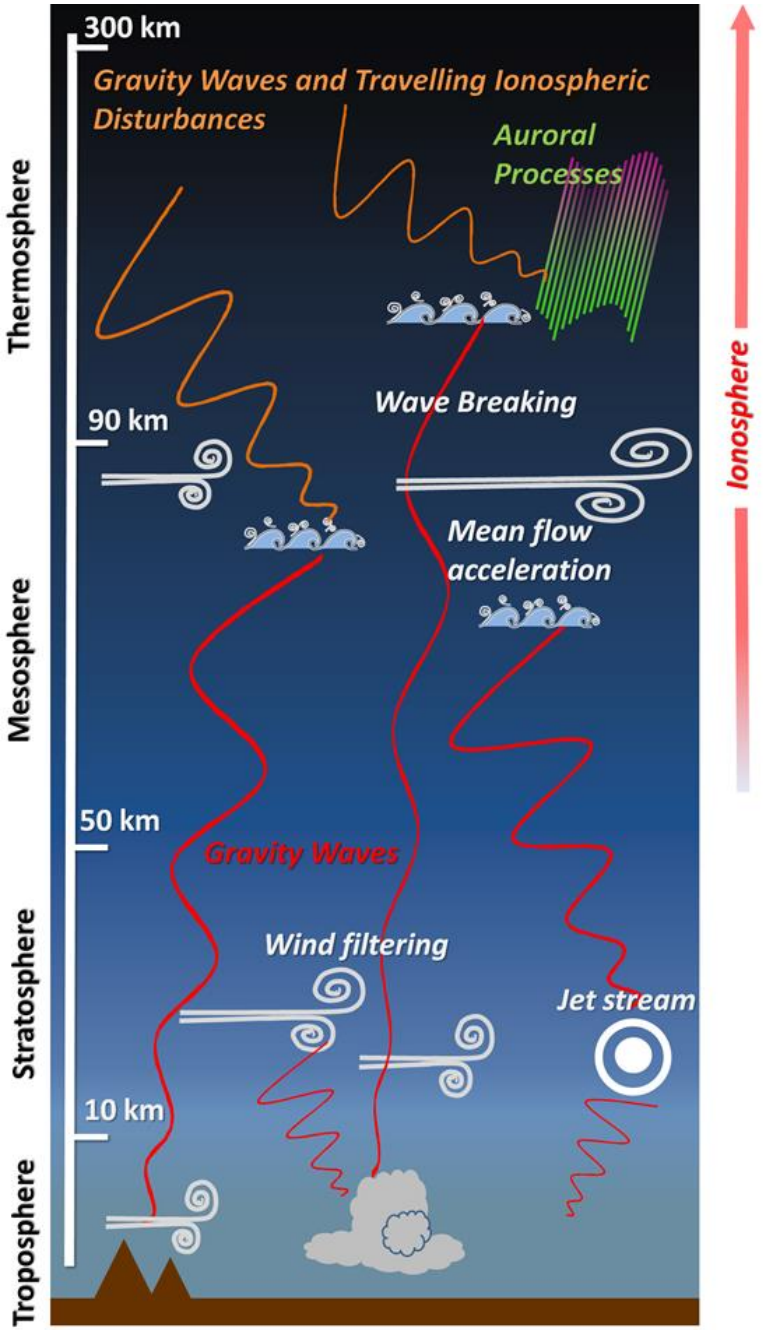
Ionospheric electrodynamic processes (e.g. atmospheric heating), usually at high latitudes

Secondary gravity waves formed by AGW breaking.

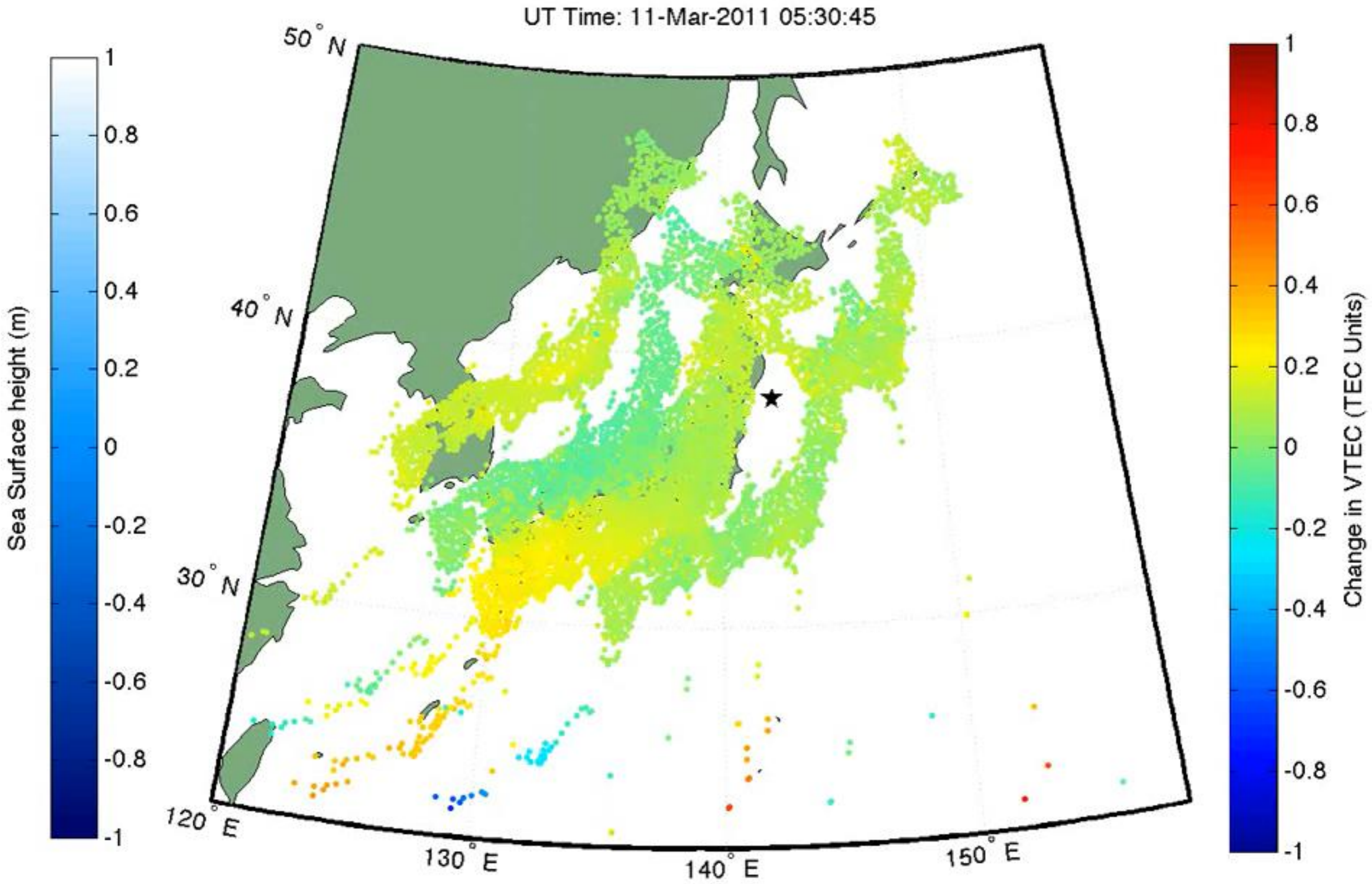
- large convective storms
- wind over orography (e.g. mountain waves)
- Instabilities such as jet streams

Large scale disturbances at ground level

- Earthquakes and Tsunamis
- Volcanic eruptions



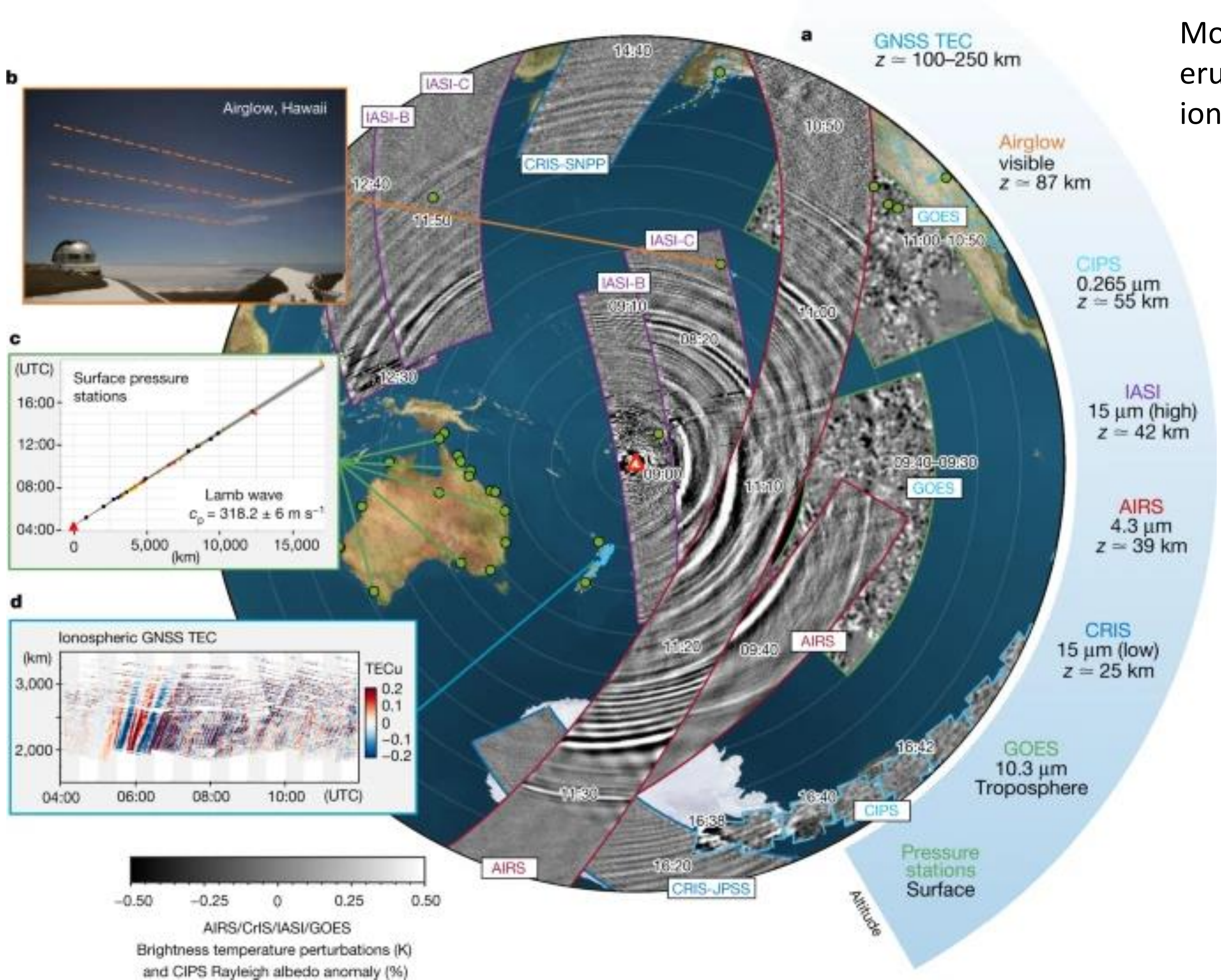
This is Total Electron Content as measured by GNSS receivers on the ground



TEC is derived from the phase delay between 2 frequencies employed by GNSS.

Can produce maps of the overall electron content and sometimes tomography to extract height information

More recently the Tonga volcanic eruption of 5th January 2022 caused ionospheric perturbations.



*Wright, C.J., Hindley, N.P., Alexander, M.J. et al. Surface-to-space atmospheric waves from Hunga Tonga–Hunga Ha'apai eruption. Nature **609**, 741–746 (2022). https://doi.org/10.1038/s41586-022-05012-5*

The High Latitude Ionosphere

>60° latitude and magnetic field lines have transitioned to being nearly vertical (especially in the polar cap)

Most highly variable region:
Solar illumination strongly varies through the year
All the compositional fun of the mid and lower latitude ionosphere PLUS space weather
Can really be split into two regions: the auroral zone and the polar cap

A new source of plasma: charged particle precipitation!***

- soft electrons with energies <1 keV (deposit in F region)
- Auroral electrons with energies from 1 – 10 keV (deposit in E and F region)
- Plasma sheet electrons with energies from > 20 keV (deposit in the E and D region)

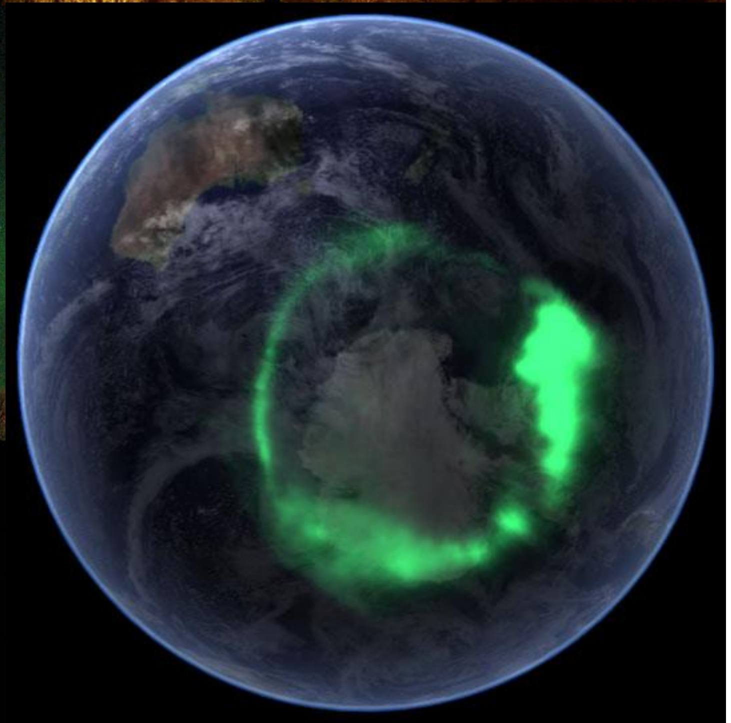
Transport becomes REALLY important in the F region:

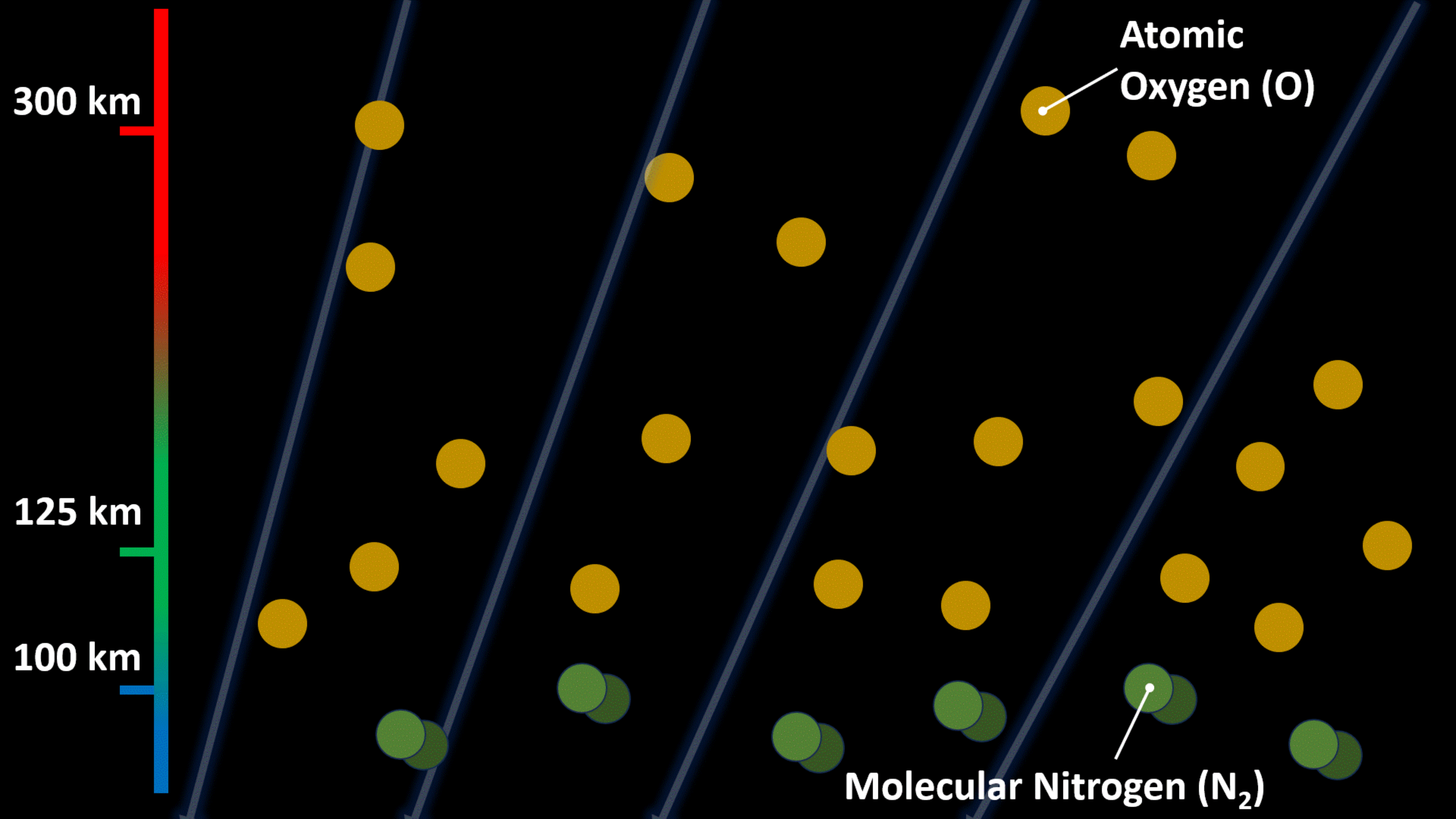
- Coupling to outer magnetosphere and magnetotail leads to strong convection
- E x B drives circulation of plasma including irregularities at different scales (e.g. plasma patches)
- Instabilities can arise in the ionospheric plasma

****As mentioned earlier precipitation can occur at low to mid latitudes as well due to connection to the radiation belts, for example* Kavanagh, A. J., Cobbett, N., & Kirsch, P. (2018). Radiation Belt slot region filling events: Sustained energetic precipitation into the mesosphere. *Journal of Geophysical Research: Space Physics*, 123, 7999–8020. https://doi.org/10.1029/2018JA025890

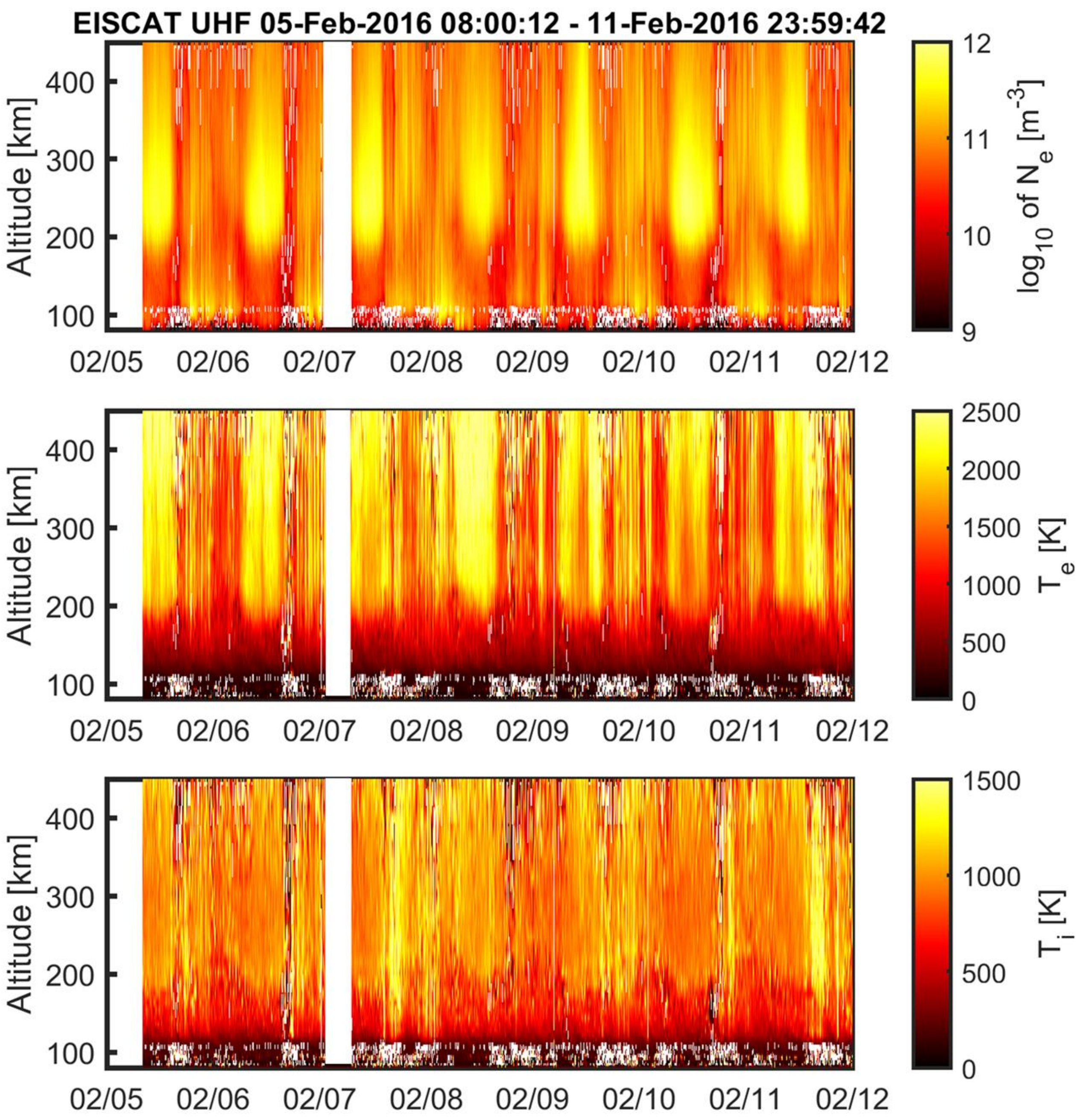




300 km
125 km
100 km
Atomic
Oxygen (O)
Molecular Nitrogen (N_2)



Data from an incoherent scatter ionospheric radar (in this case EISCAT)

Signal scatters from ion acoustic waves and the returned spectrum allows us to estimate N_e T_e T_I and the line-of-sight velocity

Can see solar ionisation

Can see periods of intense precipitation causing enhanced ionisation

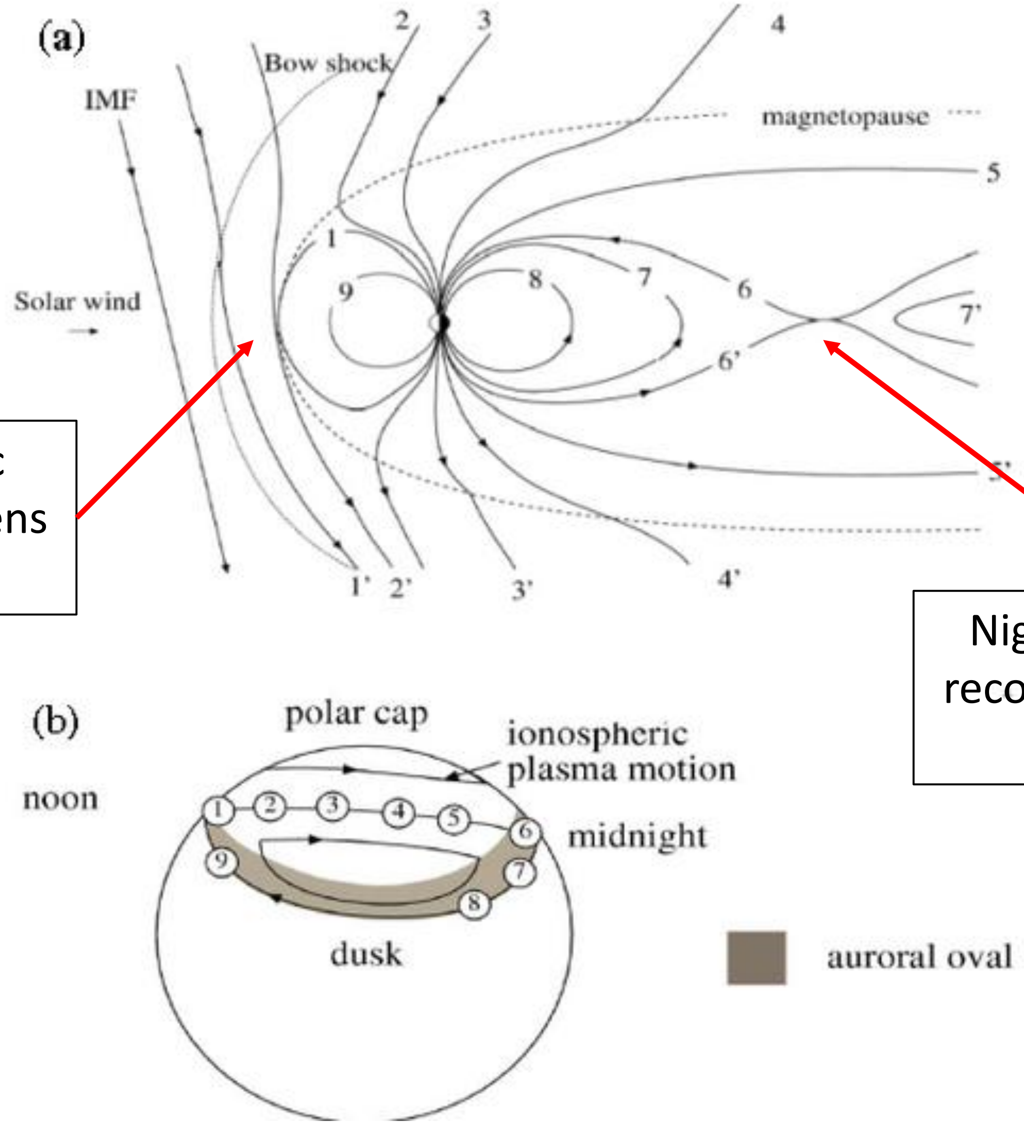
Can see periods of increased ion heating.

The latter are both strongly linked to ionospheric currents.

<u>Question:</u>

If you have charged particles precipitating into the ionosphere, what do they cause?

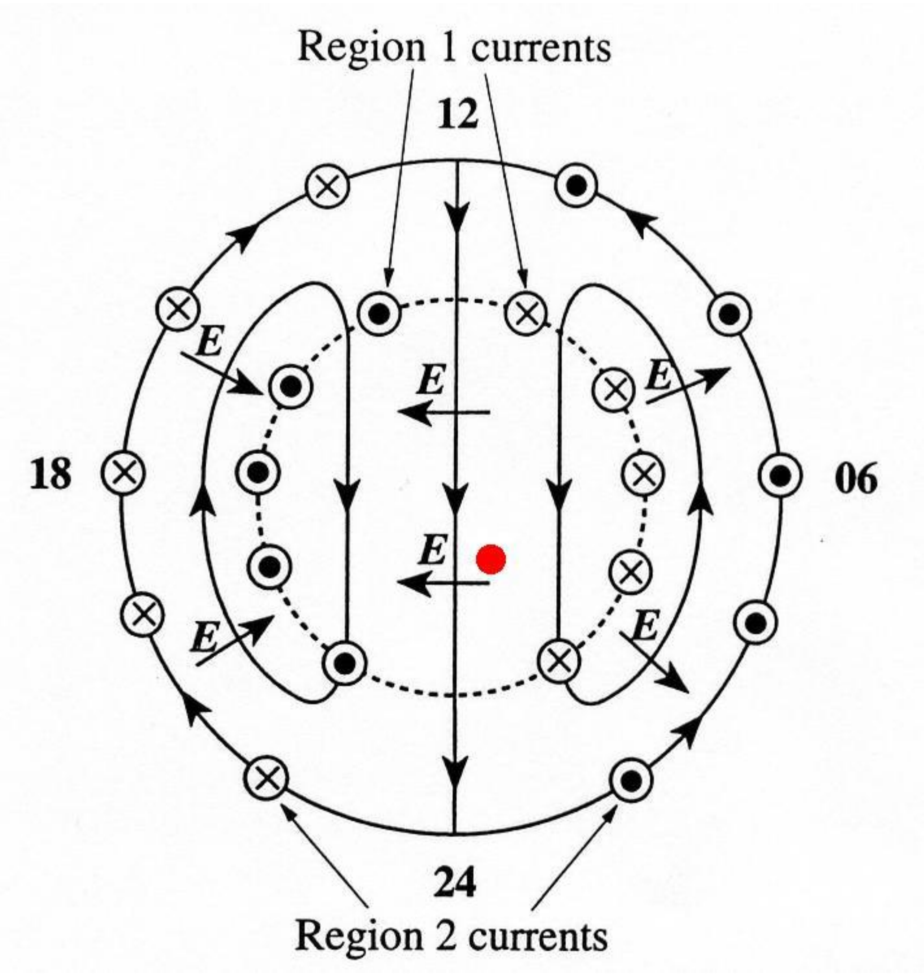
Dungey Cycle

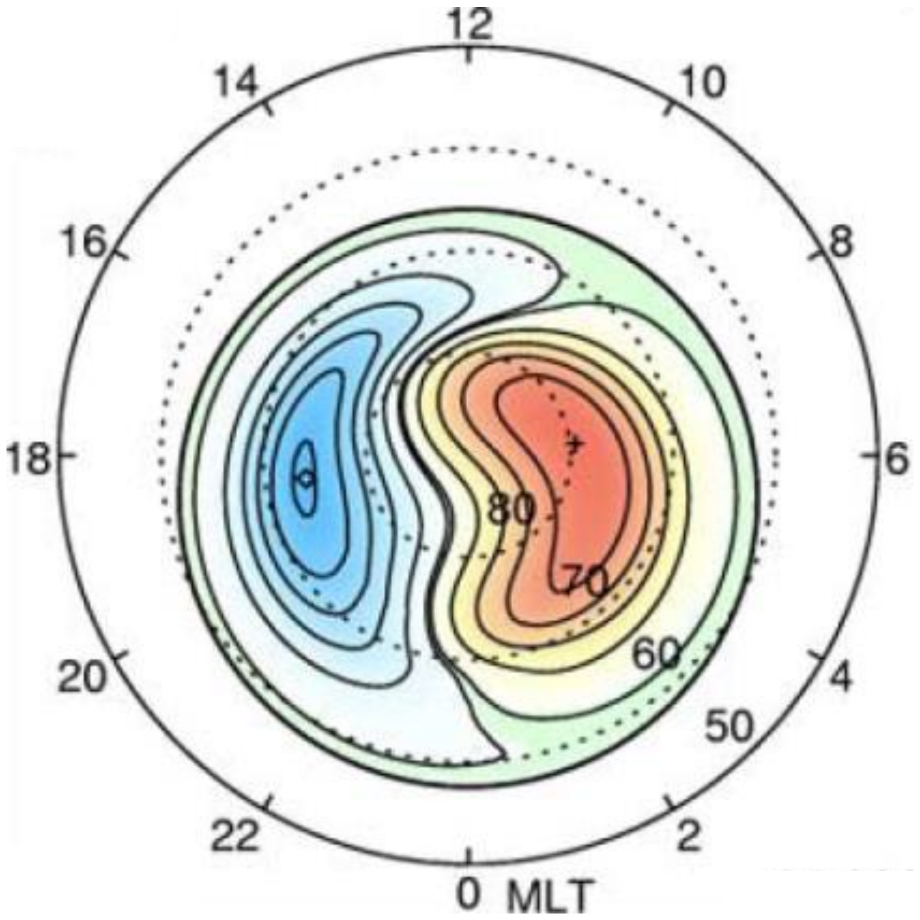


Dayside magnetic reconnection – Opens magnetic flux

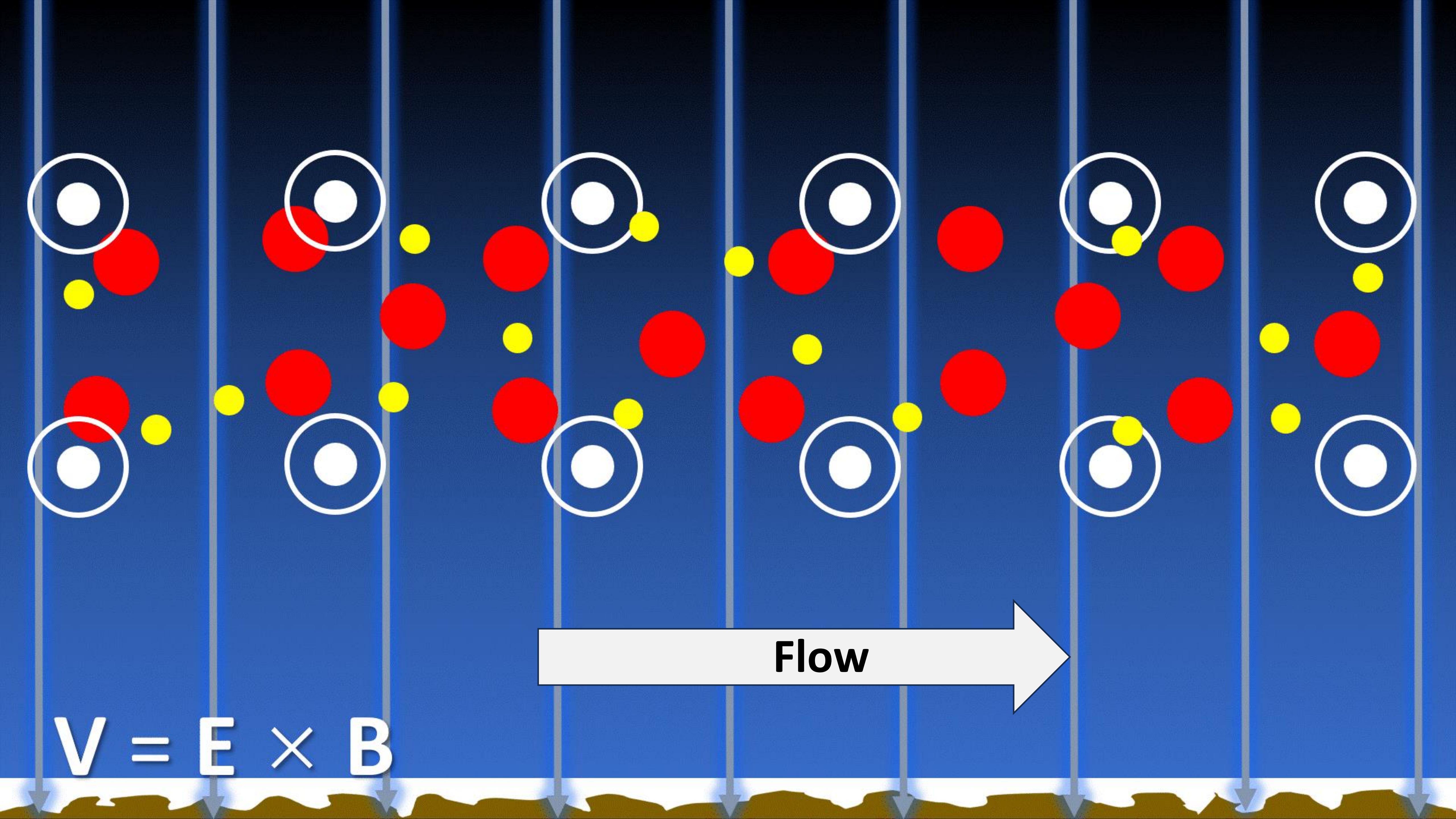
Nightside magnetic reconnection – Closes magnetic flux

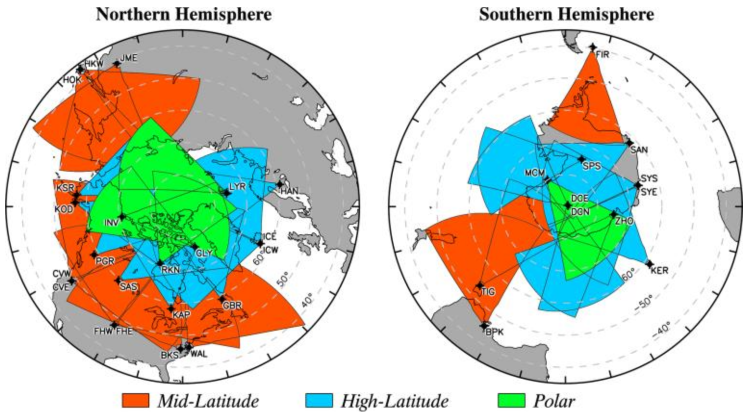
The coupling between the magnetosphere and the ionosphere results in a stirring of the ionospheric plasma

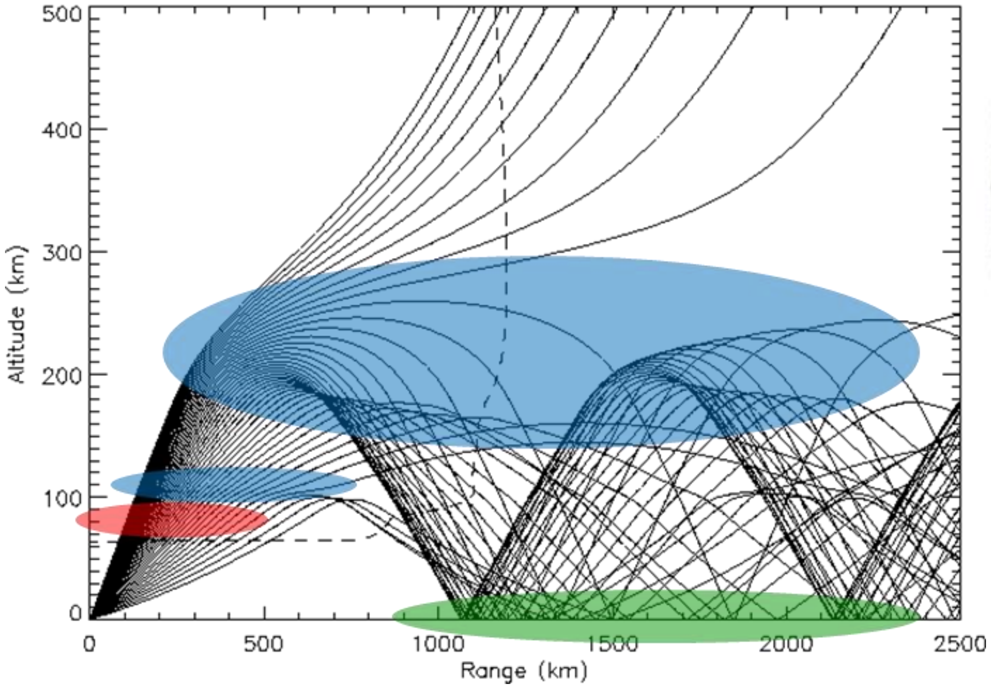


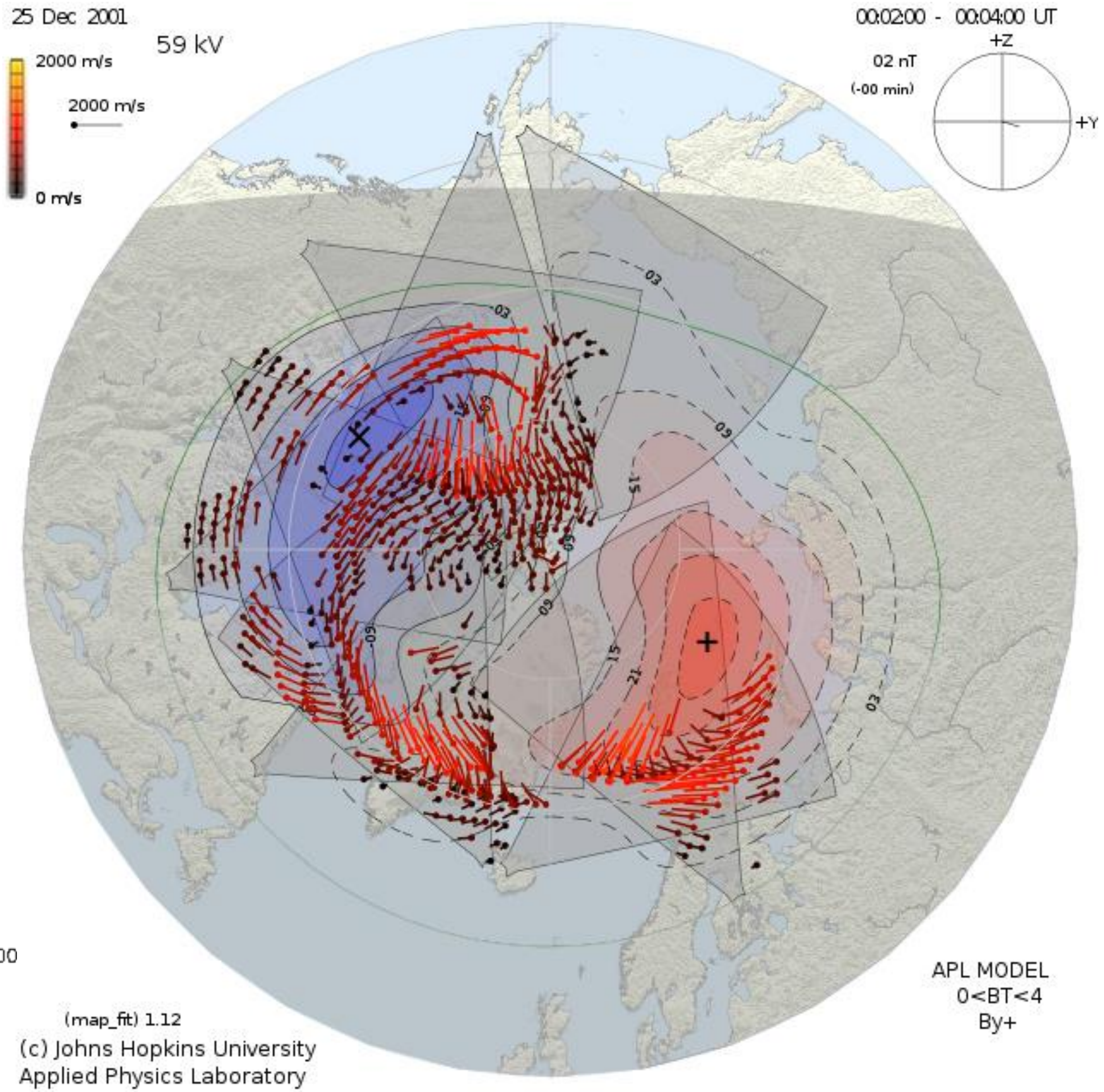


Flow streamlines are equipotential contours


Flow
V = E × B


Northern Hemisphere
Southern Hemisphere
Mid-Latitude
High-Latitude
Polar

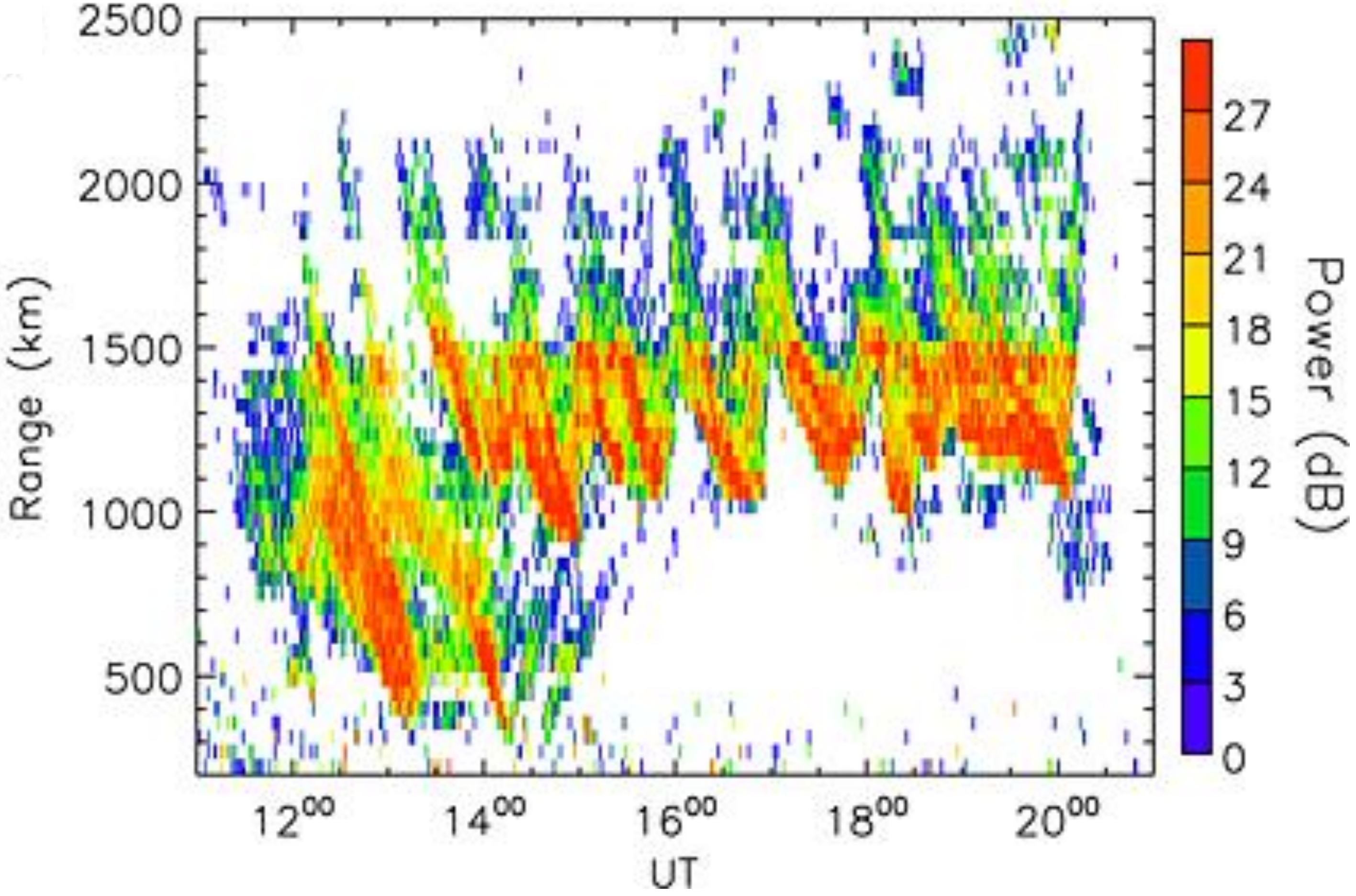

Altitude (km)
Range (km)
0
100
200
300
400
500
0
500
1000
1500
2000
2500


25 Dec 2001
59 kV
2000 m/s
2000 m/s
0 m/s
00:02:00 - 00:04:00 UT
+Z
+Y
02 nT
(-00 min)
APL MODEL
0<BT<4
By+
(map_fit) 1.12
(c) Johns Hopkins University
Applied Physics Laboratory

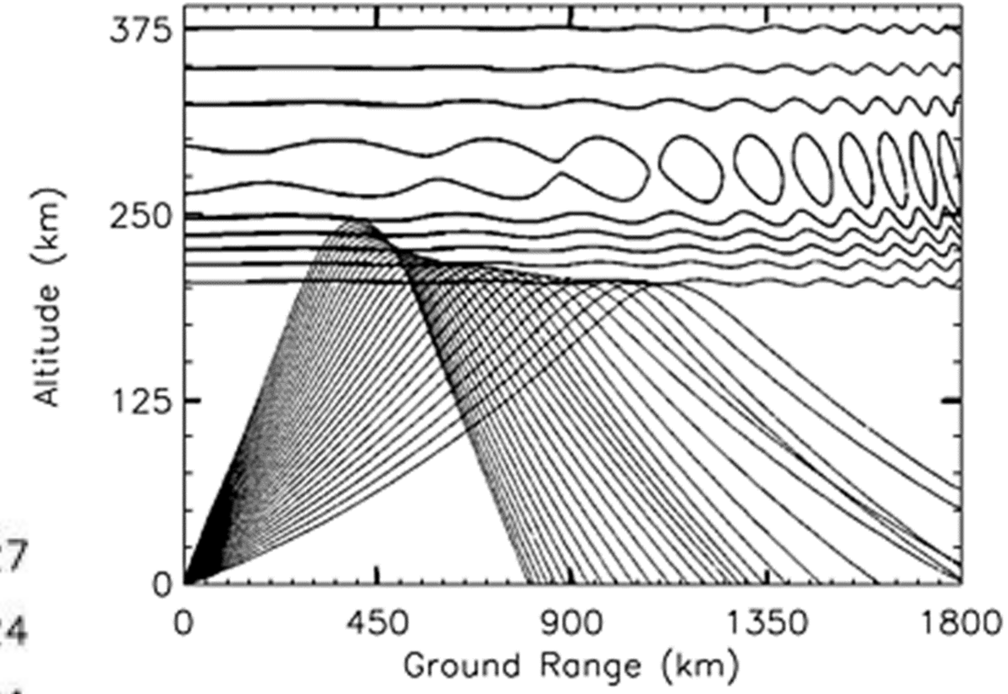
V = E × B

An aside:

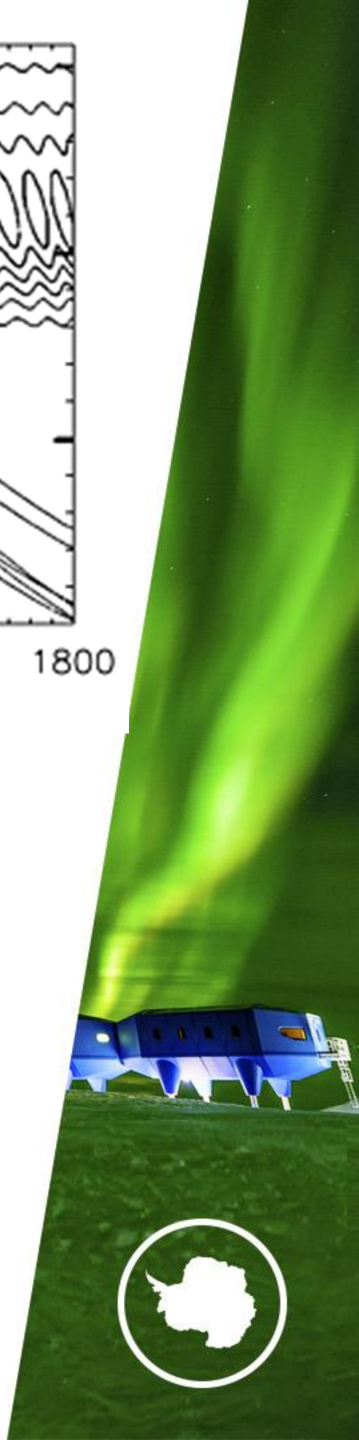
SuperDARN is an excellent means of detecting TID using the motion of the TID to alter where th ground scatter comes from.

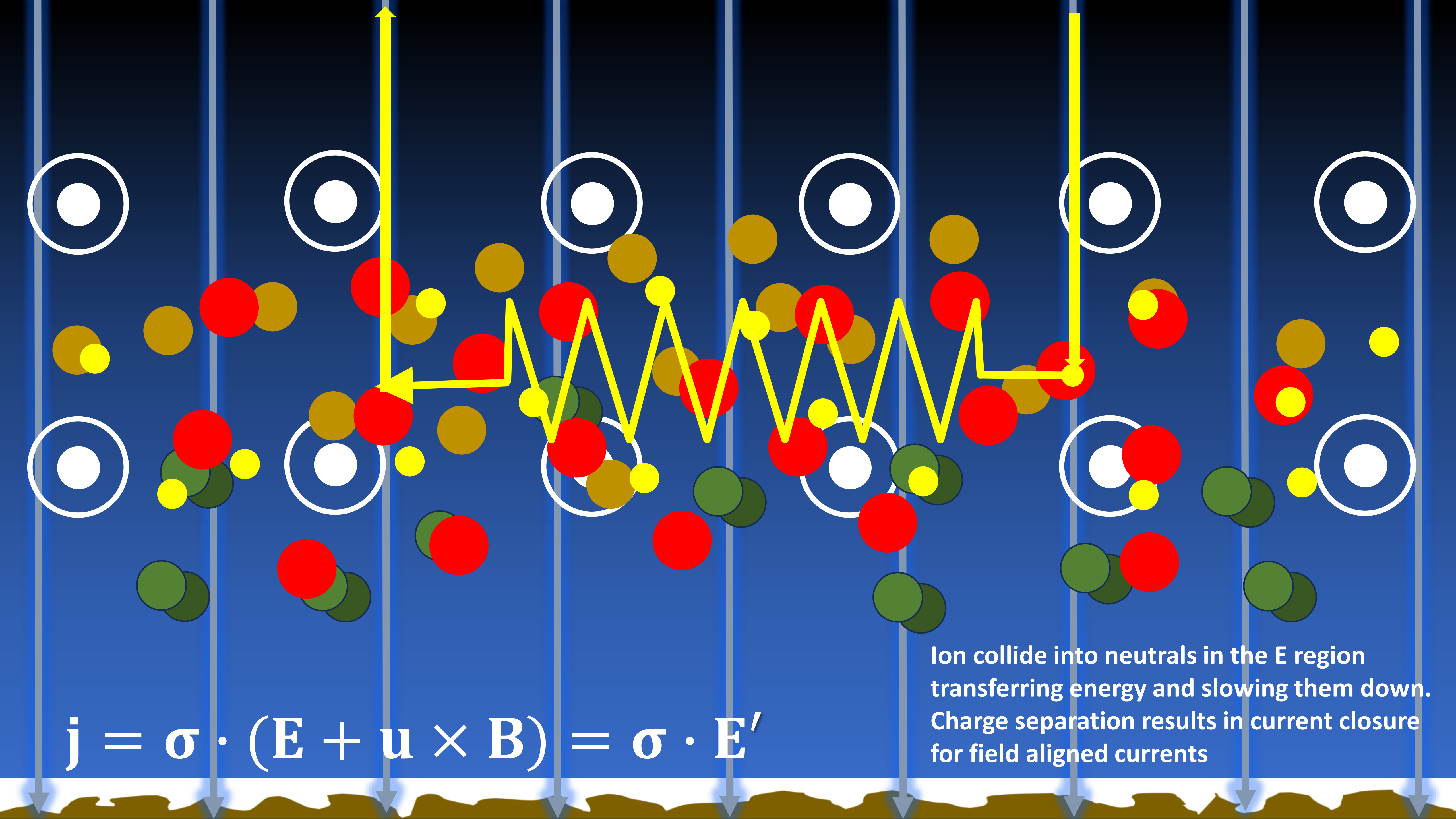


Grocott et al., 2013



A coherent scatter radar uses the ionosphere to bend the signal until it is orthogonal to field aligned irregularities from which it back scatters. The irregularities move with the convections and so the radar can measure the convection velocity




$\mathbf{j} = \boldsymbol{\sigma} \cdot (\mathbf{E} + \mathbf{u} \times \mathbf{B}) = \boldsymbol{\sigma} \cdot \mathbf{E}'$
Ion collide into neutrals in the E region transferring energy and slowing them down. Charge separation results in current closure for field aligned currents

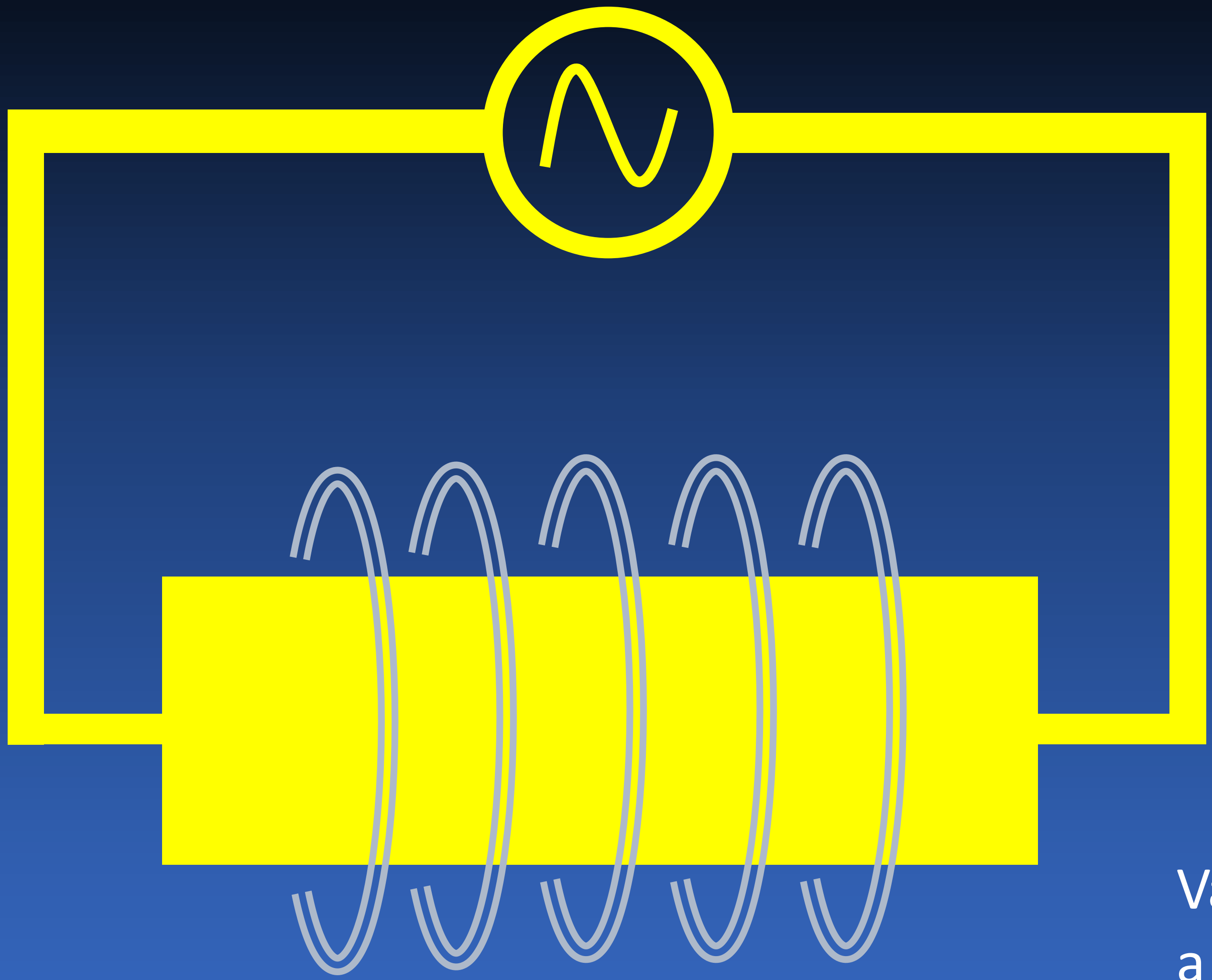
Solar wind / Space Weather = variable voltage

Upward current = auroral electrons

Downward current = protons/ions & ion upflow

Ionosphere/atmosphere = resistor

$$\mathbf{j} = \boldsymbol{\sigma} \cdot (\mathbf{E} + \mathbf{u} \times \mathbf{B}) = \boldsymbol{\sigma} \cdot \mathbf{E}'$$

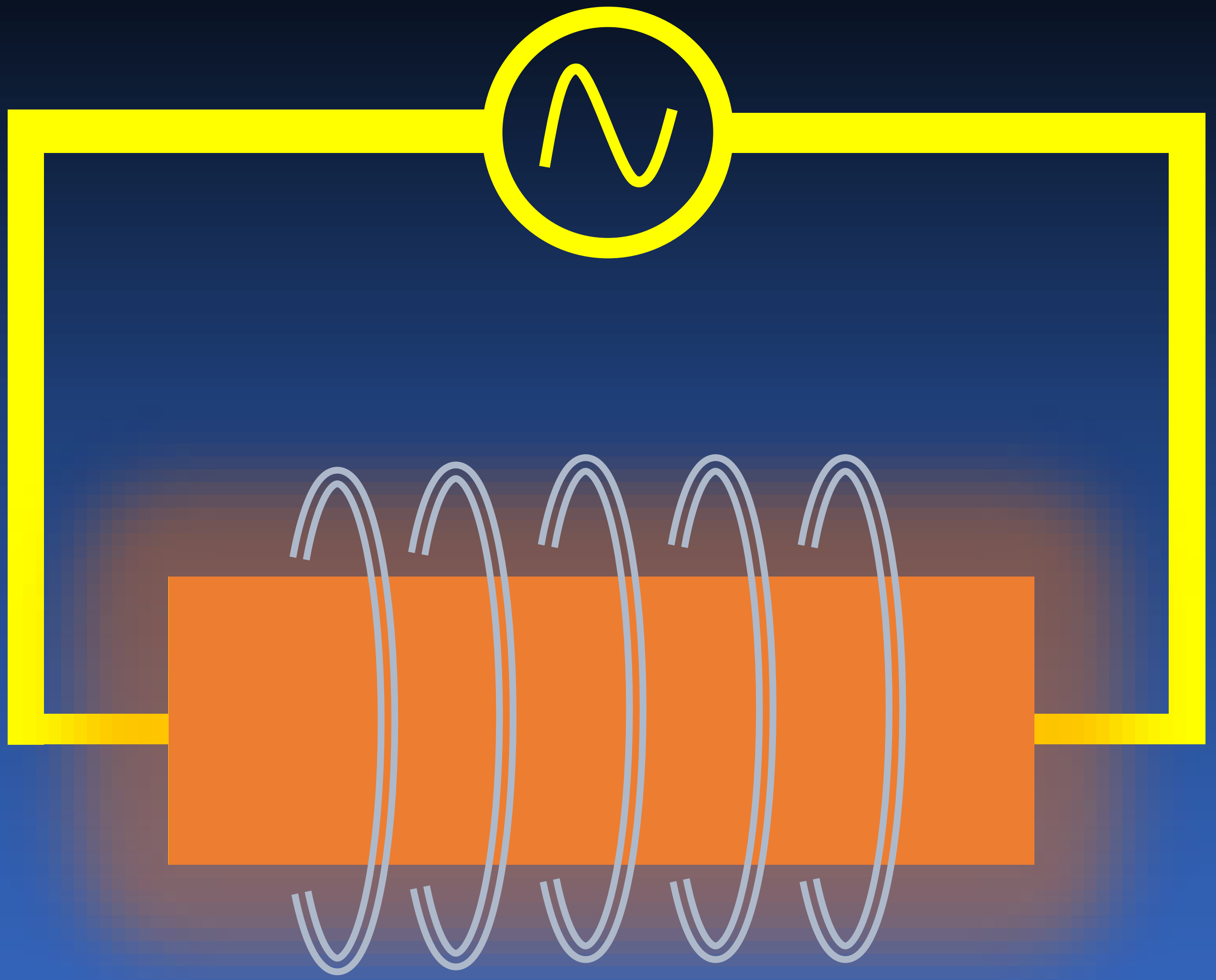


Variable current induces a magnetic field

$$\mathbf{j} = \boldsymbol{\sigma} \cdot (\mathbf{E} + \mathbf{u} \times \mathbf{B}) = \boldsymbol{\sigma} \cdot \mathbf{E}'$$

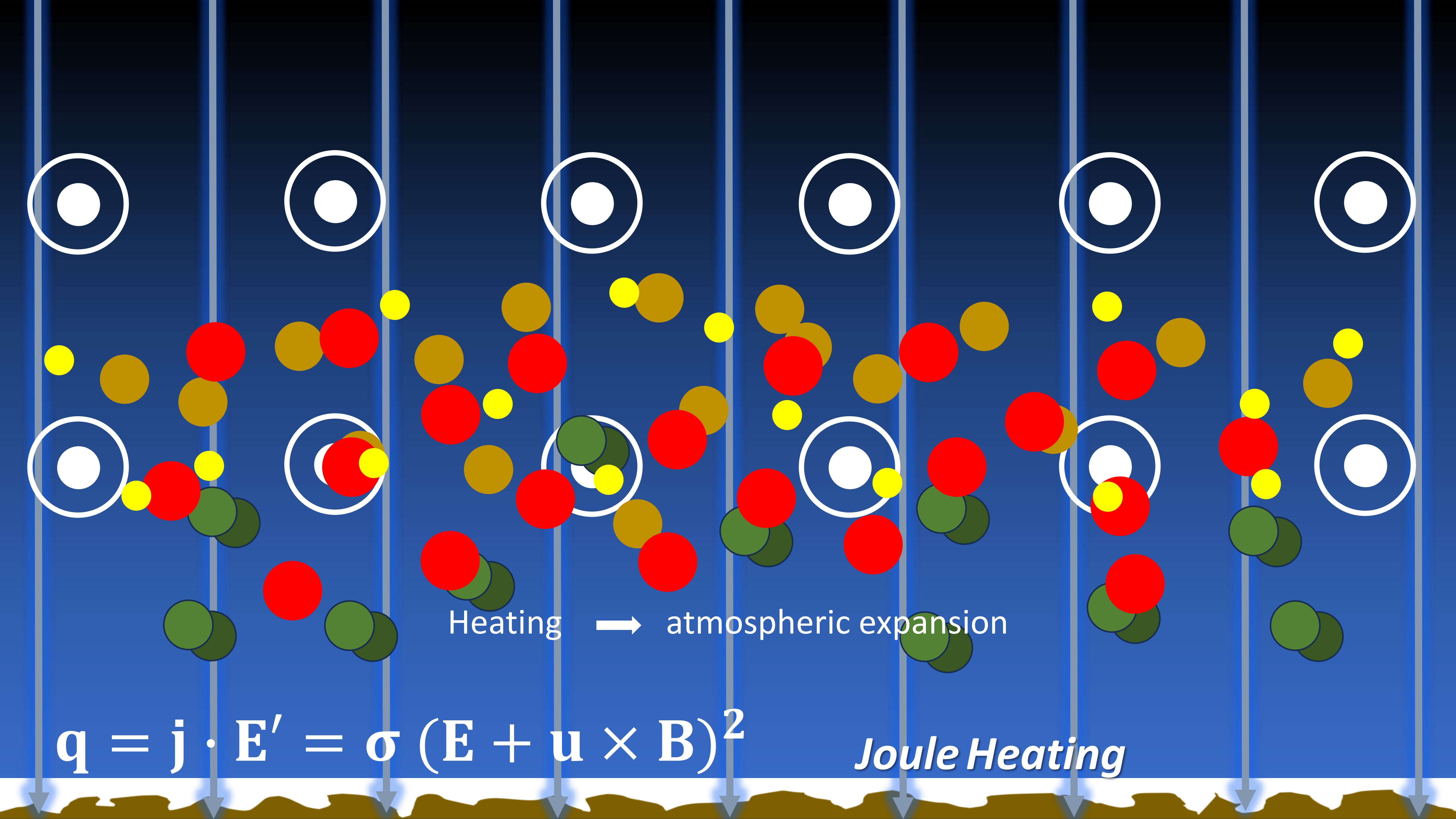
Ground Magnetometers can detect effect of ionospheric currents

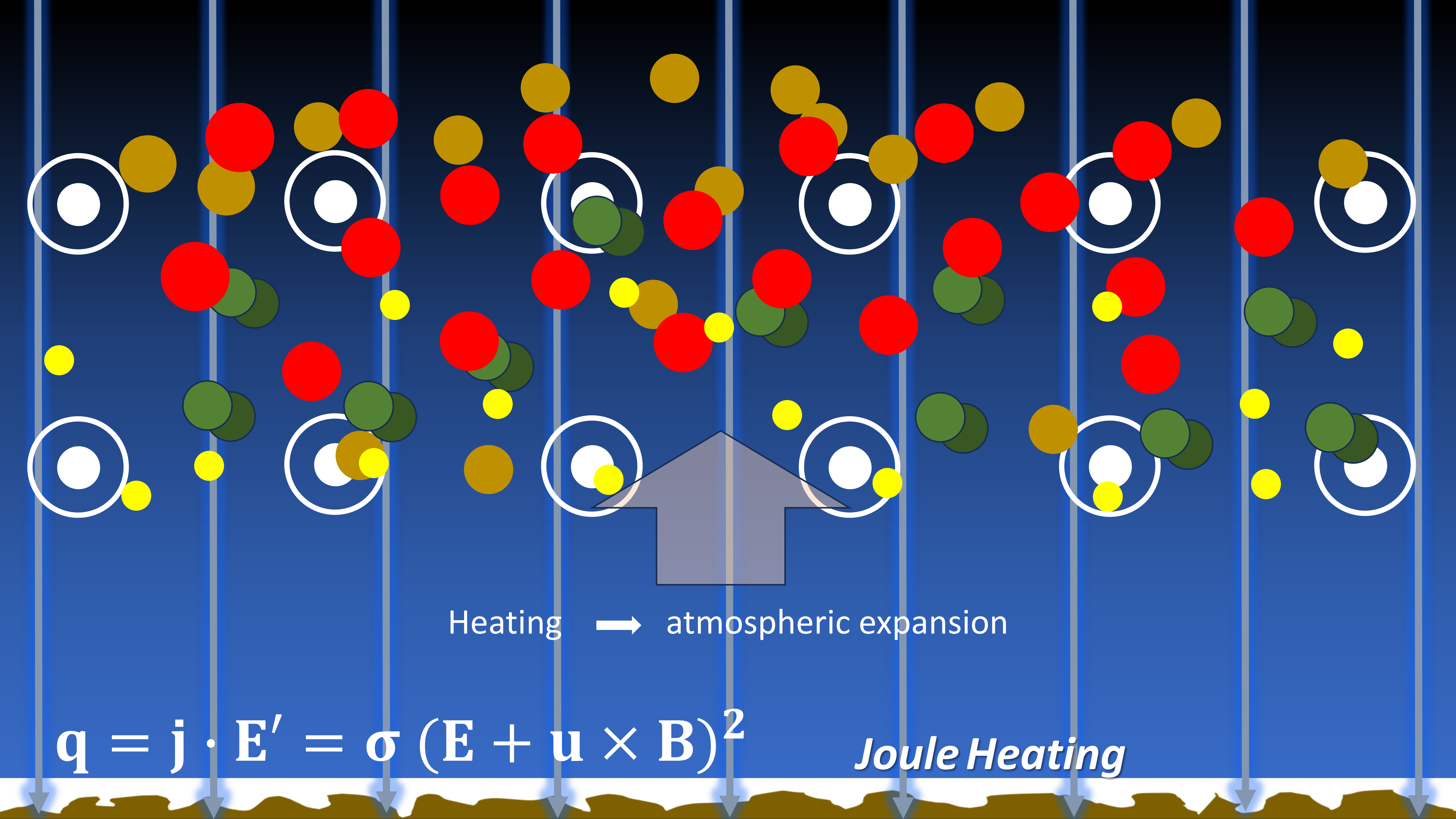
Pass a current through a resistor and it heats up



$$\mathbf{q} = \mathbf{j} \cdot \mathbf{E}' = \boldsymbol{\sigma} \left(\mathbf{E} + \mathbf{u} \times \mathbf{B}\right)^2$$

Joule Heating


Heating → atmospheric expansion
$$q = j \cdot E' = \sigma\,(E + u \times B)^2$$
Joule Heating

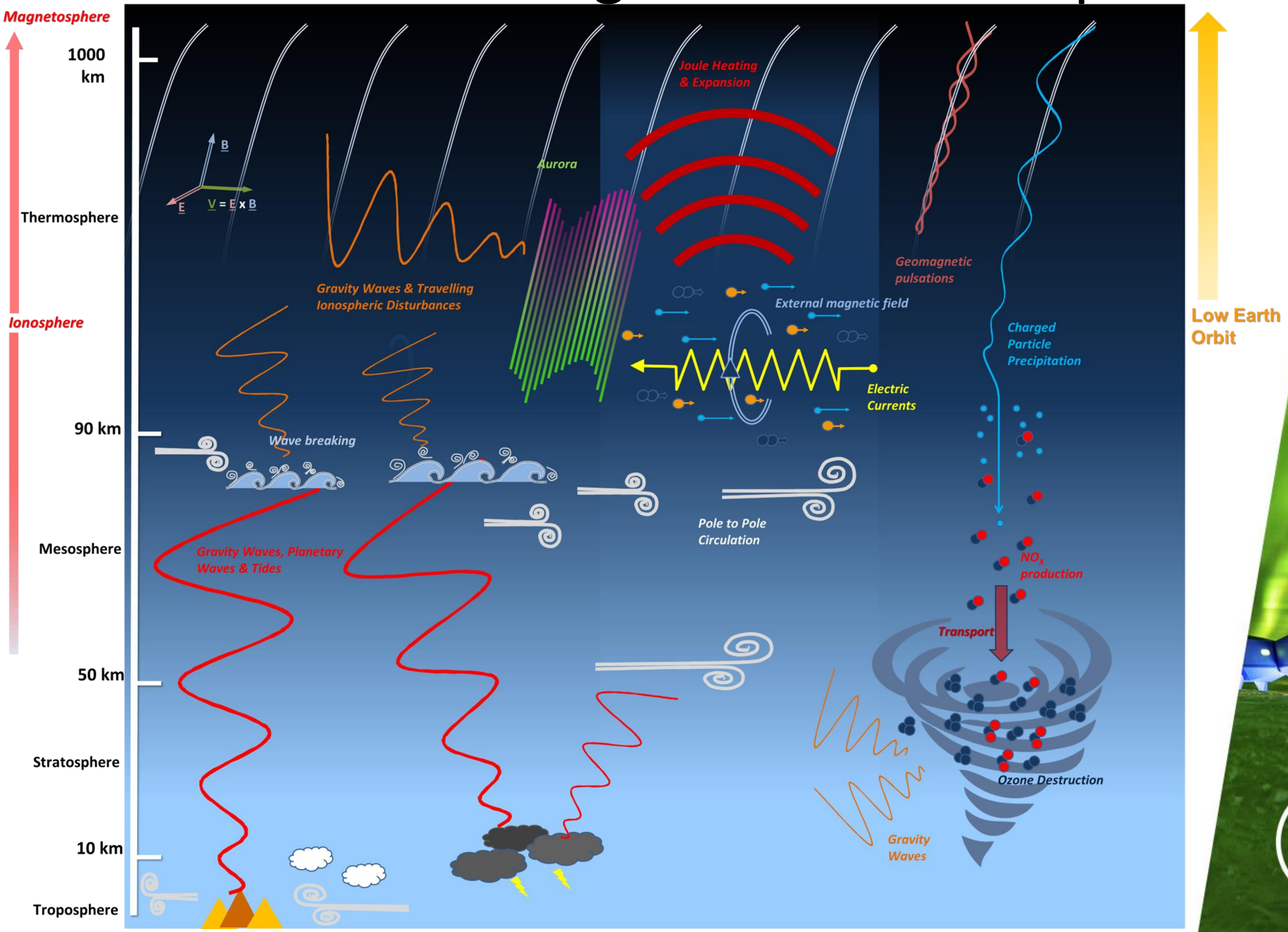


Heating → atmospheric expansion
$q = j \cdot E' = \sigma\,(E + u \times B)^2$
Joule Heating

The High Latitude Ionosphere

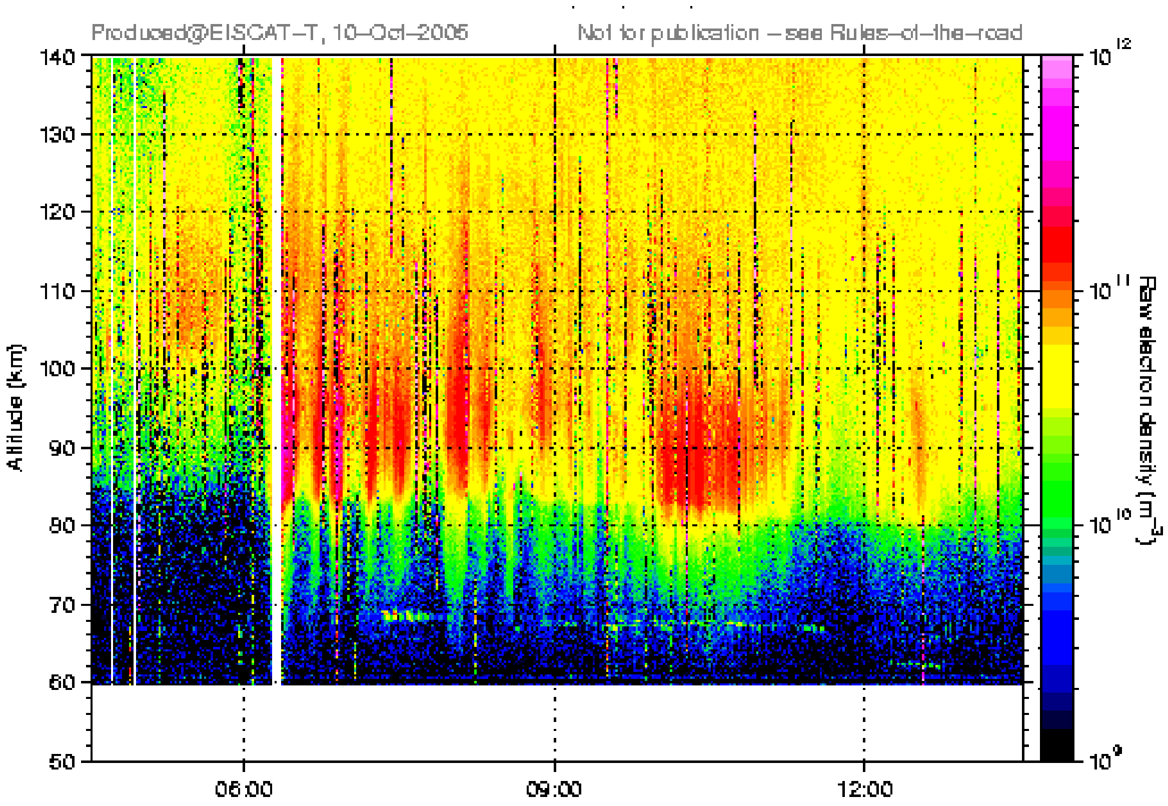
There are many important processes that occur in the high latitude ionosphere and there is not enough time to cover them all.

Many of those processes rely on the coupling with other regions of our environment.



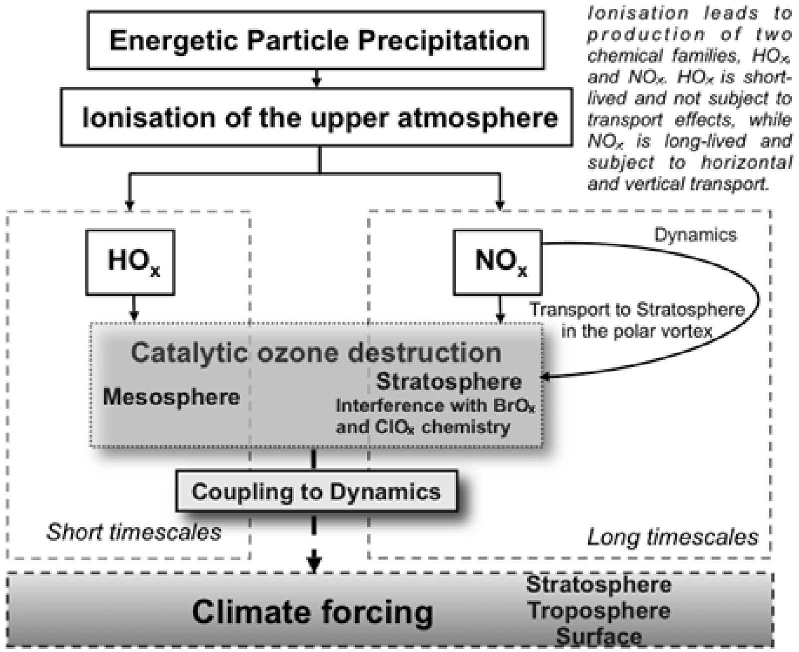
The High Latitude Ionosphere

For example:
Whereas transport is very important in the F region and in the E region (where the currents flow), it is much less important in the highly collisional D region.



Ion chemistry is VERY important.

Increased ionisation from energetic particle precipitation (10s-100s keV) leads to increases in NO_x and HO_x

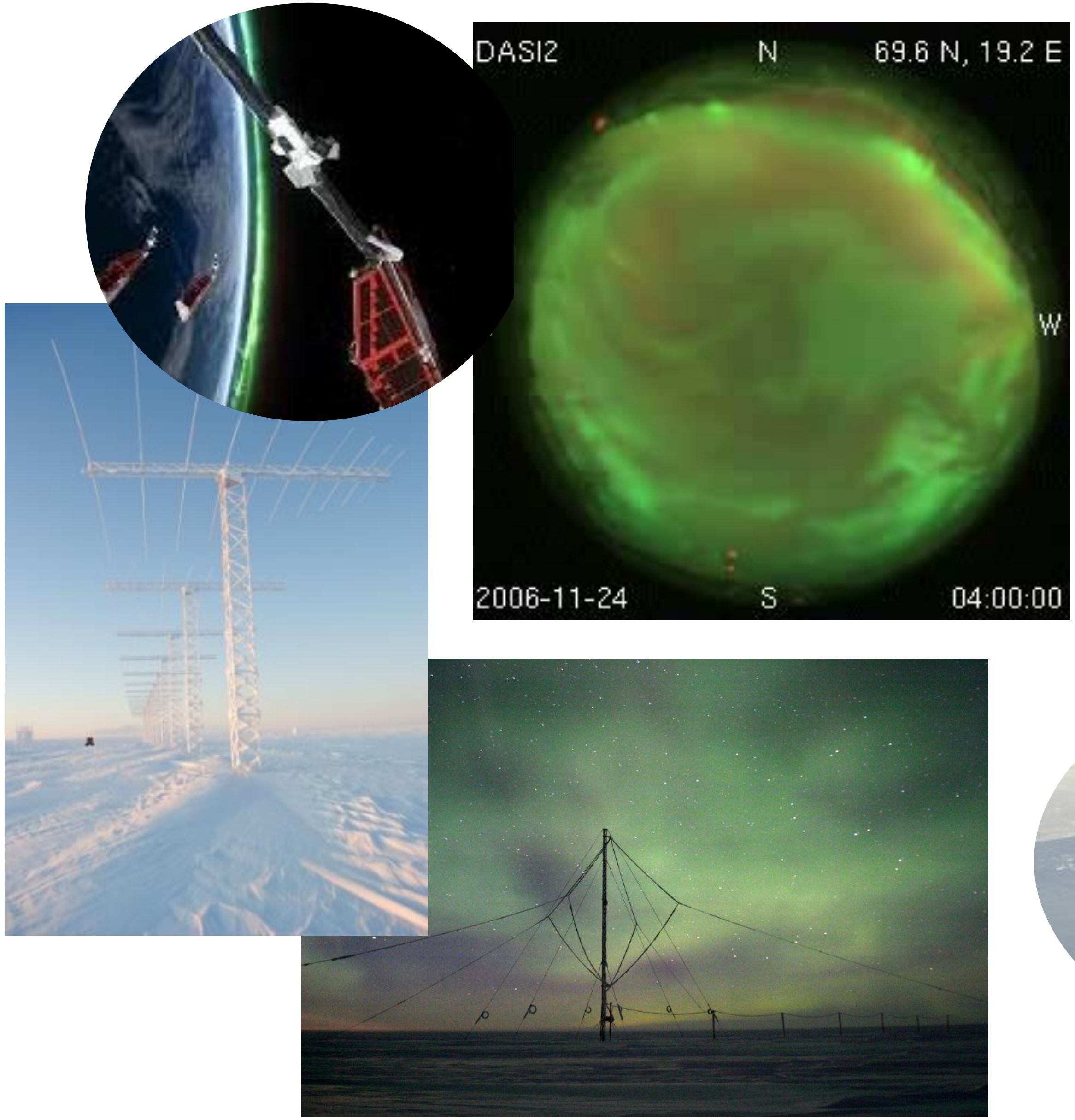


These catalytically destroy ozone and can lead to changes in the temperature and dynamics of the polar atmosphere

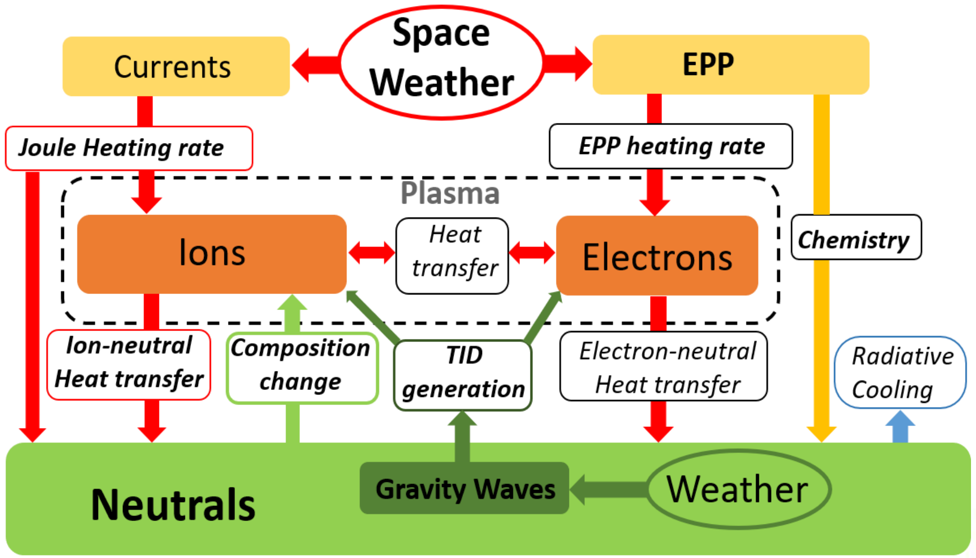
Seppälä, A., C. E. Randall, M. A. Clilverd, E. Rozanov, and C. J. Rodger (2009), Geomagnetic activity and polar surface air temperature variability, J. Geophys. Res., 114, A10312, doi:10.1029/2008JA014029. .

Measuring the Ionosphere

- It is difficult to measure large parts of the ionosphere as the atmosphere is just dense enough to cause rather short lifetimes for satellites that transit the region.
- Techniques have been developed to remote sense the region either from Space or from the Ground



What are the next big questions?



It is often better to think of the questions in terms of the whole system.

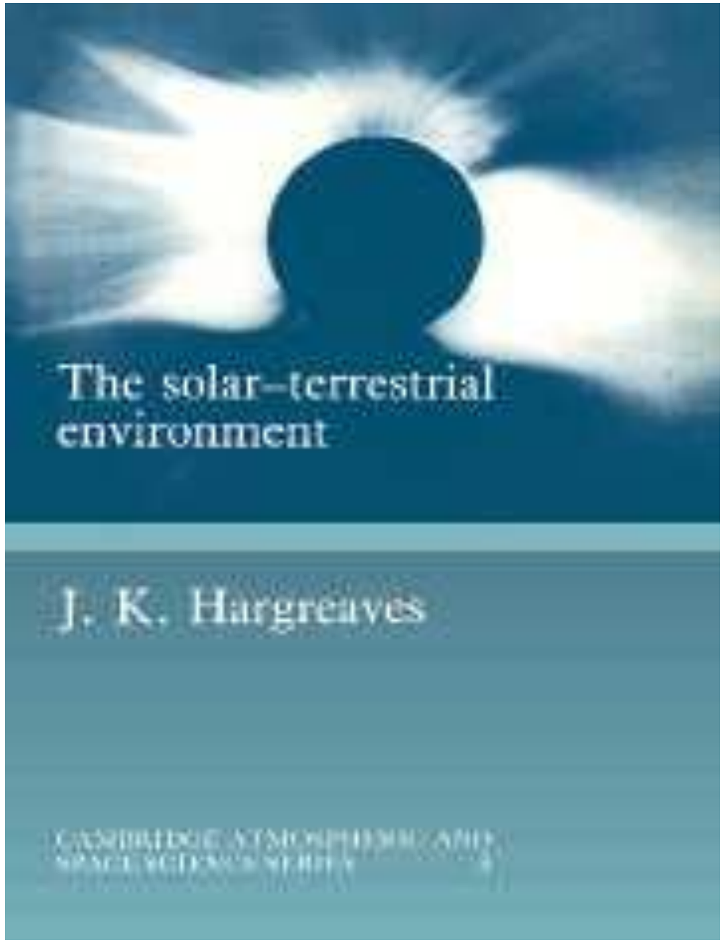
Coupling with the magnetosphere and neutral atmosphere is highly important.

Current Questions in the Ionosphere:

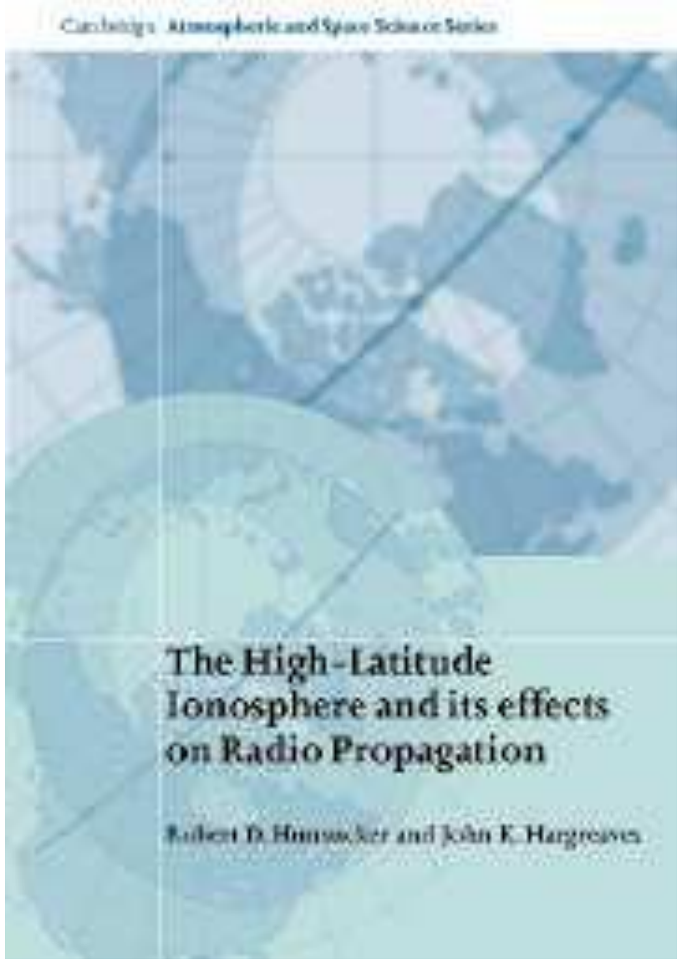
- How is climate change affecting the ionosphere and what are the implications for known processes and the energy budget?
- How does the scale size of our measurements limit our understanding? Are we missing key inputs to the system through averaging? For example, are we massively underestimating Joule heating?
- How important is the neutral-ionosphere coupling relative to the space weather driving? Is it important for generating irregularities and impacting radio propagation.
- What drive the differences in the hemispheres and are they important?
- During big geomagnetic storms, are we capturing all of the physics involved?

Ask any scientist with an interest in the ionosphere and you will get more questions.

Useful Texts to follow up:



https://www.cambridge.org/core/books/solarterrestrial-environment/76A9C5BD5FE19629EA5E9BD5DA840D5B



https://www.cambridge.org/core/books/highlatitude-ionosphere-and-its-effects-on-radio-propagation/F794D707855E6B58F6DD0618E4036862