SWARM SPACE WEATHER VARIABILITY OF IONOSPHERIC PLASMA

Swarm-VIP | Variability of Ionospheric Plasma Observed by Swarm

Swarm-VIP | Variability of Ionospheric Plasma

Swarm-VIP Team

Alan Wood¹, Luca Spogli², Yaqi Jin³, Elizabeth Donegan-Lawley¹, Jaroslav Urbar⁴, Lasse Clausen³, Lucilla Alfonsi², Golnaz Shahtahmassebi⁵, Daria Kotova³, Antonio Cicone⁶, Claudio Cesaroni², Per Høeg³, María José Brazal Aragón⁷, James Rawlings⁵ & Wojciech Miloch³

- (1) University of Birmingham, UK
- (2) Istituto Nazionale di Geofisica e Vulcanologia, Italy
- (3) University of Oslo, Norway
- (4) Institute of Atmospheric Physics CAS, Prague, Czech Republic
- (5) Nottingham Trent University, UK
- (6) Università Degli Studi Dell'Aquila, Italy
- (7) GMV Innovating Solutions, Poland





UiO **Contract University of Oslo**

JNIVERSITY OF

BIRMINGHAM





SERENE Space Environment & Radio Engineering



Swarm-VIP-Dynamic | Variability, Irregularities, and Predictive capabilities for the Dynamic ionosphere

Swarm-VIP-Dynamic Team

Alan Wood¹, Gareth Dorrian¹, Daria Kotova², Luca Spogli³, Yaqi Jin², Jaroslav Urbar⁴, Eelco Doornbos⁵, Lucilla Alfonsi², Kasper van Dam⁵, Mainul Hoque⁶ & Wojciech Miloch³

University of Birmingham, UK
 University of Oslo, Norway
 Istituto Nazionale di Geofisica e Vulcanologia, Italy
 Institute of Atmospheric Physics CAS, Prague, Czech Republic
 KNMI, Royal Netherlands Meteorological Institute, the Netherlands
 German Aerospace Center (DLR), Germany





UiO **: University of Oslo**



Koninklijk Nederlands Meteorologisch Instituut Ministerie van Infrastructuur en Waterstaat





CAS

INSTITUTE OF ATMOSPHERIC PHYSICS





Swarm-VIP | Variability of Ionospheric Plasma

Swarm-VIP Project:

Investigate the variability in ionospheric plasma on multiple spatial scales

Generalised Linear Modelling:

Determines the relative importance of the driving processes & which combination of driving processes best represent the observed variability

Create a model of the variability of ionospheric plasma





UiO **Contensity of Oslo**

INIVERSITY OF

BIRMINGHAM



Space Environment & Radio Engineering



Generalised Linear Modelling | Method

Linear Model

$$E(y) = \beta_0 + \beta_1 \cdot x_1$$

Multivariate Linear Model

$$E(y) = \beta_0 + \beta_1 \cdot x_1 + \dots + \beta_n \cdot x_n$$

 $g(E(y)) \models \beta_0 + \beta_1 \cdot x_1 + \dots + \beta_n \cdot x_n$

Generalised Linear Model

The dependent variable does not have to follow a normal distribution

Equation may have a different form

- E(y) is the expected value of the dependent variable y
- $x_1 \cdots x_n$ are the independent, or explanatory, variables
- $\beta_1 \cdots \beta_n$ are the parameter estimates for the model

Dependent variable y

Explanatory variables $x_1 \cdots x_n$

To be predicted May influence y



Helio-geophysical proxies



Swarm Observations | Variability of Ionospheric Plasma

Measure of the variability of the electron density: Level 2 data product IPDxIRR_2F from IPIR

|Grad_Ne@100km|: The electron density gradient in a running window calculated via linear regression over 27 data points for the 2 Hz electron density data

[Grad_Ne@50km]: The electron density gradient in a running window calculated via linear regression over 13 data points for the 2 Hz electron density data

[Grad_Ne@20km]: The electron density gradient in a running window calculated via linear regression over 5 data points for the 2 Hz electron density data

Electron Density: Measured by the Langmuir probe

IPIR Index: The product of the rate of change density index in 10 seconds (RODI10*s*) and the standard deviation of the electron density in a running window of 10 seconds ($A(n_e)_{10s}$):

 $IPIR_{ix} = RODI10s \cdot A(n_e)_{10s}$



Swarm Observations | Variability of Ionospheric Plasma

Measure of the variability of the electron density: Level 2 data product IPDxIRR_2F from IPIR

[Grad_Ne@100km]: The electron density gradient in a running window calculated via linear regression over 27 data points for the 2 Hz electron density data

[Grad_Ne@50km]: The electron density gradient in a running window calculated via linear regression over 13 data points for the 2 Hz electron density data

[Grad_Ne@20km]: The electron density gradient in a running window calculated via linear regression over 5 data points for the 2 Hz electron density data

Electron Density: Measured by the Langmuir probe

IPIR Index: The product of the rate of change density index in 10 seconds (RODI10*s*) and the standard deviation of the electron density in a running window of 10 seconds ($A(n_e)_{10s}$):

 $IPIR_{ix} = RODI10s \cdot A(n_e)_{10s}$

Latitudinal region: Modelling polar, auroral, mid-latitude and equatorial regions separately **Two years of data from Swarm A,** covering 16th July 2014 – 15th July 2015 & 1st January – 31st December 2017

Model Development | Choice of Distribution



Model Development | Choice of Distribution

Example: |Grad_Ne@100km| in the polar region: Gamma distribution



Heliogeophysical proxies

Interplanetary Magnetic Field (IMF): Bt, Bx, By, |By|, Bz and clock angle. Mean and standard deviation measured over a 2 hour interval, starting 2 hours before the observation of the dependent variable.

Solar wind: Mean and standard deviation of bulk speed, density, pressure and Interplanetary Electric Field (IEF) measured over a 2 hour interval, starting 2 hours before the observation of the dependent variable.

Solar wind coupling functions: Newell (left) and E_{LYA} (right). Mean and standard deviation of these measured over a 2 hour interval, starting 2 hours before the observation of the dependent variable.

$$v^{4/3} \cdot B_T^{2/3} \cdot \sin^{8/3}\left(\frac{\theta_c}{2}\right) \qquad \qquad v B_T^{\frac{1}{2}} \sin^2\left(\frac{\theta_c}{2}\right)$$

v is the solar wind velocity, B_T is the magnitude of the IMF and θ_c is the clock angle. **Geomagnetic Activity:** AE, AL, AU, ASY-D, ASY-H, SYM-D, SYM-H, Kp and Ap **Solar Activity:** F10.7 cm solar flux (daily, 27 day median and 81 day median) **Other:** Polar cap index (North), GLAT, MLAT, MLT and SZA



Generalised Linear Modelling | Method



Multi term model: Which combination of heliogeophysical proxies, and hence which processes, best explain the variability of ionospheric plasma observed?

- Add the heliogeophysical proxy which best explains the variability to the model (the most statistically-significant proxy)
- Add the heliogeophysical proxy which best explains the remaining variability (the next most statistically significant, excluding any proxy which is correlated with any term already in the model by more than [0.25])
- Continue until no more proxies make a significant difference (no more are statistically significant at the 5% level)



Example Model: Polar model of |*Grad_Ne*@100km|

$$\sqrt[3]{|Grad_Ne@100km|} = \exp\left(\begin{array}{c}\beta_0 + \beta_1 \cdot F107_{81} + \beta_2 \cdot SZA + \beta_3 \cdot fDOY + \beta_4 \cdot Kp \\ +\beta_5 \cdot |MLAT| + \beta_6 \cdot B_x + \beta_7 \cdot SW_Den + \beta_8 \cdot SYM_D\end{array}\right)$$

- F107₈₁ 81 day average of the F10.7cm solar flux, centred on the day to be updated
- SZA Solar Zenith Angle
- fDOY A sine function based on the day of year, going from -1 at midwinter to +1 at midsummer

Kp Planetary K-index

- |MLAT| Absolute value of magnetic latitude (in degrees)
- BxThe average value of the x-component of the Interplanetary Magnetic Field (in nT), across a two hourwindow, starting two hours before the observation to be updated
- SW_Den As B_x , but for the solar wind proton number density (in # cc^{-1})
- SYM_D The longitudinally symmetric disturbances to the terrestrial magnetic field perpendicular to the dipole axis



Database

- 1st January 2014 15th July 2014:
- 16th July 2014 15th July 2015:
- 16th July 2015 31st December 2016:
- 1st January 2017 31st December 2017:
- 1st January 2018 28th February 2018:

Optimisation and Evaluation Training Optimisation and Evaluation Training

- Optimisation and Evaluation
- Odd DOYs used for optimisation and even DOYs used for evaluation



Swarm-VIP Model | Optimisation & Evaluation

Optimisation: Refit models. Identify any terms which are no-longer statistically-significant at the 5% level. Remove these.

AIC is based on the trade off between model performance and model complexity

 $AIC = -2 \cdot ln(L) - 2 \cdot k$

L is the log likelihood, k is the number of parameters

Does the addition of successive terms to the model reduce the AIC? If not, then remove terms that increase the AIC

Do not remove terms based on the AIC if there is no longer a measure of LAT, MLT/ST/SZA, Solar Activity & geomagnetic activity/solar wind. The items in each category are:

- LAT: LAT, MLAT
- MLT/ST/SZA: MLT, LT, SZA
- Solar Activity: F10.7cm Solar Radio Flux, Sunspot Number
- Geomagnetic activity/solar wind: Any IMF, Any measure of the solar wind or a coupling function, any geomagnetic index



Optimisation Example: Polar model of |*Grad_Ne*@100km|

$$\sqrt[3]{|Grad_Ne@100km|} = \exp\left(\begin{array}{c}\beta_0 + \beta_1 \cdot F107_{81} + \beta_2 \cdot SZA + \beta_3 \cdot fDOY + \beta_4 \cdot Kp \\ +\beta_5 \cdot |MLAT| + \beta_6 \cdot B_x + \beta_7 \cdot SW_Den + \beta_8 \cdot SYM_D\end{array}\right)$$

Re-fit model and remove any terms which are no-longer significant: Lose two terms

$$\sqrt[3]{|Grad_Ne@100km|} = \exp\left(\begin{array}{c}\beta_0 + \beta_1 \cdot F107_{81} + \beta_2 \cdot SZA + \beta_3 \cdot fDOY + \beta_4 \cdot Kp \\ + \beta_5 \cdot |MLAT| + \beta_7 \cdot SW_Den\end{array}\right)$$

Is this level of complexity justified? Lose another term

$$\sqrt[3]{|Grad_Ne@100km|} = \exp\left(\begin{array}{c}\beta_0 + \beta_1 \cdot F107_{81} + \beta_2 \cdot SZA + \beta_3 \cdot fDOY + \beta_4 \cdot Kp \\ + \beta_5 \cdot |MLAT|\end{array}\right)$$



Swarm-VIP Model | Evaluation

Accuracy: RMSE

$$RMSE = \sqrt{\frac{1}{N-d} \sum_{i=1}^{N} (M_i - O_i)^2}$$

Expressed on a relative scale (divide RMSE by median of observations). If rRMSE=1 then the differences between observations and predictions are the same size as the predictions.

Bias: Mean Error

$$ME = \overline{M} - \overline{O}$$

Positive: Model consistently overpredicts. Negative: Model consistently underpredicts

Precision: Ratio of the standard deviations

$$P_{\sigma,\text{ratio}} = \frac{\sigma_M}{\sigma_O}$$

If this is greater than 1, then the average standard deviation of the model values is greater than of the observed values. There is too much spread (noise) in the model values. If it is less than 1, then there is not enough spread in the model values (model is overfitted).

Association: Pearson Linear Correlation Coefficient

Model values are denoted by M, with individual values with the number set listed as M_i . Observational values are given the variable O, with individual data points called out by O_i . The total number of pairs in the data-model set is N.

Liemohn et al., RMSE is not enough: Guidelines to robust data-model comparisons for magnetospheric physics, JASTP, 218, 105624, 2021

$$R = \frac{\sum \left(O_i - \overline{O}\right) \left(M_i - \overline{M}\right)}{\sqrt{\sum \left(O_i - \overline{O}\right)^2 \sum \left(M_i - \overline{M}\right)^2}}$$



Goodness of fit statistics

Model		Goodness of fit				
Region	Dependent Variable	rRMSE	rME	Precision	Correlation	
	GradNe@100km	0.47	0.00	0.37	0.36	
	GradNe@50km	0.49	0.00	0.37	0.35	
Polar	GradNe@20km	0.58	0.00	0.31	0.30	
	IPIR Index	0.24	0.00	0.46	0.45	
	Ne	0.16	0.00	0.76	0.75	
	GradNe@100km	0.47	0.00	0.31	0.31	
	GradNe@50km	0.18	0.00	0.28	0.28	
Auroral	GradNe@20km	0.98	0.01	0.26	0.24	
	IPIR Index	0.24	0.00	0.37	0.36	
	Ne	0.15	0.00	0.76	0.75	
	GradNe@100km	0.19	0.00	0.24	0.23	
	GradNe@50km	0.19	0.00	0.22	0.21	
Mid Latitude	GradNe@20km	0.98	0.00	0.20	0.20	
	IPIR Index	0.30	0.00	0.33	0.34	
	Ne	0.16	0.00	0.73	0.68	
Equatorial	GradNe@100km	0.43	-0.05	0.57	0.30	
	GradNe@50km	0.21	-0.02	0.56	0.29	
	GradNe@20km	1.16	-0.12	0.61	0.28	
	IPIR Index	0.34	-0.04	0.48	0.19	
	Ne	0.16	-0.03	0.83	0.53	

rRMSE: Substantially less
than 1 in most cases
rME: No significant bias
Precision: Not enough
spread in model values
Correlation: Between
0.20 and 0.75



TIEGCM Comparisons | Along-track comparisons

TIE-GCM simulations: Model predictions on a grid of:

- 2.5° latitude
- 2.5° longitude
- Pressure levels separated by 0.25 scale heights
- one hour

TIEGCM captures some, but not all, of the ionospheric variability





TIEGCM Comparisons | Along-track comparisons

There are occasions when TIEGCM represents the Swarm observations well, such as when the ionosphere is dominated by photoionisation When the ionosphere is variable, TIEGCM does not always capture that variability TIEGCM does not always capture ionospheric structures even during quiet conditions





TIEGCM Comparisons | Along-Track Comparisons

Predictions and observations of electron density compared for four week long case studies

Model	Region	Goodness of fit					
MOdel		rRMSE	rME	Precison	Correl.		
Swarm-VIP	Polar	0.16	0.02	0.55	0.65		
	Auroral	0.15	0.01	0.70	0.62		
	Mid	0.16	0.00	0.76	0.48		
	Equatorial	0.14	-0.01	0.72	0.43		
TIEGCM	Polar	0.24	-0.11	0.37	0.22		
	Auroral	0.24	-0.14	0.42	0.22		
	Mid	0.17	-0.03	0.46	0.38		
	Equatorial	0.13	0.07	0.81	0.69		

 Dates & orbits
 4th - 10th Sept. 2017: 10-22 LT, Major geomagnetic storm
 26th April - 2nd May 2017: 10-22 LT
 13th - 20th August 2017: 00-12 LT
 23th - 29th October 2017: 06-18 LT

- **Correlations:** Swarm VIP models show a moderate improvement over TIE-GCM in the polar, auroral and mid-latitude sectors.
- **Correlations:** TIE-GCM shows a moderate improvement over Swarm VIP models in the equatorial sector
- Bias: Less bias in Swarm VIP models

Spogli, L., et al., (2024). Statistical Models of the Variability of Plasma in the Topside Ionosphere: 2: Performance Assessment, J. Space Weather Space Clim., swsc230023. doi: 10.1051/swsc/2024003

Jin et al., Performance Assessment for Statistical Models of the Ionospheric Variability in the Topside Ionosphere, poster, this meeting



Goodness of fit statistics

	Mo	odel Goodness of fit					
Regio	Region Dependent Variable		rRMSE	rME	Precision	Correlation	
		GradNe@100km	0.47	0.00	0.37	0.36	
		GradNe@50km	0.49	0.00	0.37	0.35	
Polar	-	GradNe@20km	0.58	0.00	0.31	0.30	
Auror Mid Latit	Auror • Can only capture the deterministic component • The explanatory variables are proxies for processes, not the processes themselves • Is there a missing process: No thermospheric term in models						
IPIR Index			0.30	0.00	0.33	0.34	
	Ne 0.16 0.00 0.73 0.68						
GradNe@100km 0.43 -0.05 0.57 0.3							
		GradNe@50km	0.21	-0.02	0.56	0.29	
Equato	rial	GradNe@20km	1.16	-0.12	0.61	0.28	
IPIR Index 0.34 -0.04 0.48 0.						0.19	
		Ne	0.16	-0.03	0.83	0.53	

rRMSE: Substantially less than 1 in most cases
rME: No significant bias
Precision: Not enough spread in model values
Correlation: Between
0.20 and 0.75



Swarm VIP Dynamic | Ionosphere-Thermosphere Coupling

- **DNSxPOD_2** Level 2 data product which estimates the neutral density based on Precise Orbit Determination (POD)
- Cadence of 30 seconds, but ~20 minutes of data used for each data point
- ACCxCAL2 Level 2 data product which includes the along track accelerations
- Cadence of 1 second
- **DNSxACC_2_** Level 2 data product which estimates the neutral density based on the accelerations
- Cadence of 10 seconds



Swarm-VIP-Dynamic | Ionosphere-Thermosphere Coupling

Polar model of |*Grad_Ne*@100km|

Original model: No thermospheric information

$$\sqrt[3]{|Grad_Ne@100km|} = \exp\left(\begin{array}{c}\beta_0 + \beta_1 \cdot F10.7_{81} + \beta_2 \cdot SZA + \beta_3 \cdot fDOY + \beta_4 \cdot Kp \\ + \beta_5 \cdot |MLAT|\end{array}\right)$$

Densities inferred from Precise Orbit Determination: DNSxPOD_2_

$$\sqrt[3]{|Grad_Ne@100km|} = \exp\left(\begin{array}{c} \beta_0 + \beta_1 \cdot Density_{POD} + \beta_2 \cdot fDOY + \beta_3 \cdot SZA \\ +\beta_4 \cdot Clock + \beta_5 \cdot A_p + \beta_6 \cdot B_x + \beta_7 \cdot Clock_{sd} \end{array}\right)$$

Along track accelerations: ACCxCAL2

$$\sqrt[3]{|Grad_Ne@100km|} = \exp\left(\begin{array}{c} \beta_0 + \beta_1 \cdot Acc_X + \beta_2 \cdot SZA + \beta_3 \cdot |MLAT| + \beta_4 \cdot fDOY \\ + \beta_5 \cdot SW_Vel + \beta_6 \cdot B_ysd \end{array}\right)$$

Densities inferred from accelerometers: DNSxACC_2_

$$\sqrt[3]{|Grad_Ne@100km|} = \exp\left(\begin{array}{c} \beta_0 + \beta_1 \cdot Density_{acc} + \beta_2 \cdot |MLAT| + \beta_3 \cdot SZA \\ + \beta_4 \cdot fDOY + \beta_5 \cdot SW_Vel + \beta_6 \cdot B_ysd \end{array}\right)$$

Swarm-VIP-Dynamic | Ionosphere-Thermosphere Coupling

Ap

Example Model: Polar model of |*Grad_Ne@100km*|

Grad_Ne@100km in the polar region		Version				SZA
		No density	DNS_POD	ACC	DNS_ACC	fDC
	SZA					
	fDOY					
es	MLAT					M
iabl	F10.7 ₈₁					F10
var	B _x					
ory	B _y sd					R
nat	Clock					D _x
pla	Clock _{sd}					
EX	SW_Vel					
	Кр					
	Ар					Clo

SZA	Solar Zenith Angle
fDOY	A sine function based on the day of year, going from -1 at midwinter to +1 at midsummer
MLAT	Abs. value of magnetic latitude (degrees)
F107 ₈₁	81 day average of the F10.7cm solar flux, centred on the day to be updated
B _x	The average value of the x-component of the Interplanetary Magnetic Field (in nT), across a two hour window, starting two hours before the observation
Clock	As Bx for clock angle
Clock _{sd}	As Clock, but standard deviation
SW_Den	As B _x , but for the solar wind velocity (in km s ⁻¹)
Кр	Planetary K-index

Geomagnetic index

PIARILITY OF IONOSPHERIC PLASM

Swarm-VIP-Dynamic | Ionosphere-Thermosphere Coupling

Example Model: Polar model of |*Grad_Ne@100km*|



Change in goodness of fit statistics is small

- Need to include lags
- Some issues with data selection, based upon quality flags
- Interhemispherical and longitudinal effects still to be considered
- Swarm-VIP-Dynamic includes substantial amounts of work on model development

Model		Goodness of fit			
Dependent variable	ent variable Version		rME	Precision	Correlation
	No thermosphere	0.47	0.00	0.37	0.36
Grad_Ne@100km	DNS_POD	0.50	0.00	0.44	0.36
in the polar region	ACC	0.49	0.00	0.34	0.31
	DNS_ACC	0.48	0.00	0.39	0.34



Statistical models allow reasonable predictions of ionospheric plasma, and the variability in this plasma, to be predicted

The models have successes and they do not show significant bias

The models do not capture all of the ionospheric variability; they cannot capture stochastic variations and there is at least one missing driver, the thermosphere

Performance Assessment: Jin et al., Performance Assessment for Statistical Models of the Ionospheric Variability in the Topside Ionosphere, poster, this meeting

Swarm-VIP-Dynamic

Will address limitations of Swarm-VIP models, plus many other questions Miloch et al., 15:30 today

Models from the Swarm-VIP project

Wood, A. G. et al. (2024). Statistical Models of the Variability of Plasma in the Topside Ionosphere: 1. Development and Optimisation, J. Space Weather Space Clim., in press, swsc230022.

Spogli, L., et al., (2024). Statistical Models of the Variability of Plasma in the Topside Ionosphere: 2: Performance Assessment, J. Space Weather Space Clim., swsc230023. doi: 10.1051/swsc/2024003

