Thermospheric Neutral Species Reduced Order Probabilistic Emulator: Dimensionality Reduction

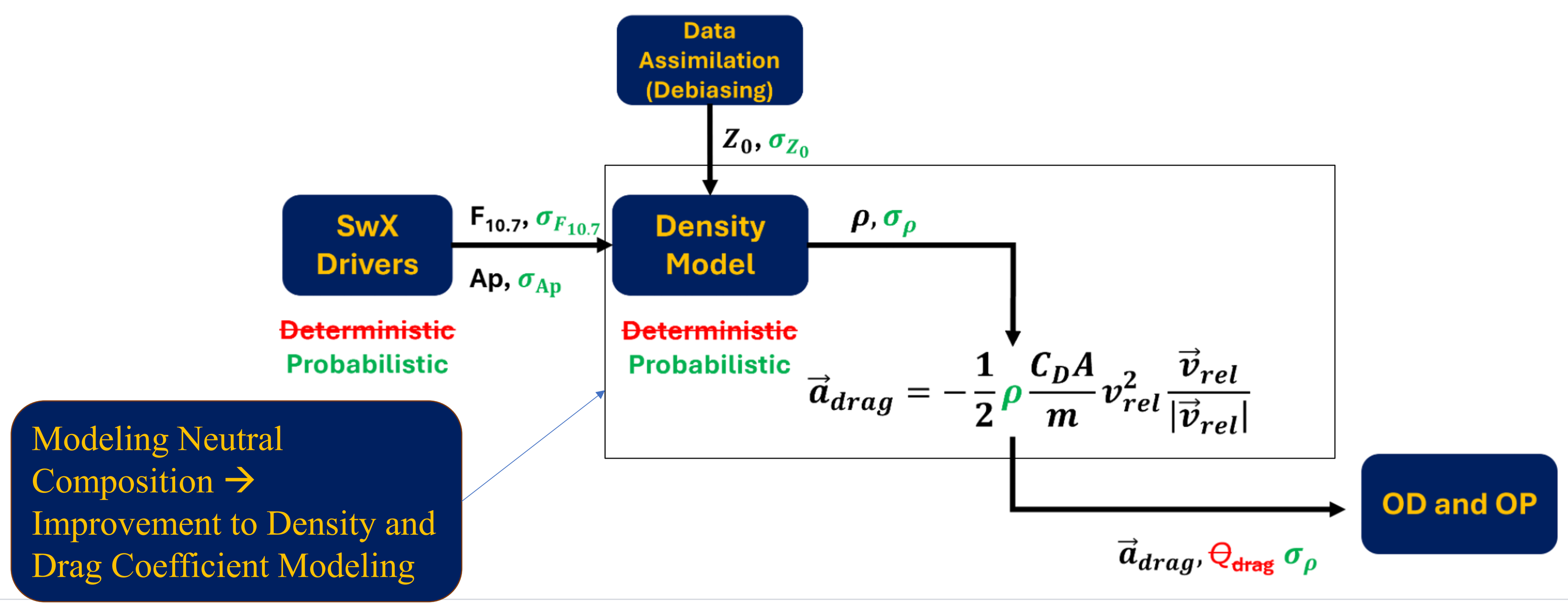
Nathaniel Michek[1] · Piyush Mehta[1]

[1] ASSIST Lab, West Virginia University

Background

Previous Development

- **Reduced Order Probabilistic Emulator (ROPE)[1-3]**
- **TIE-GCM ROPE → 10,000x speedup**



Benefits of Modeling Neutral Composition

TIE-GCM ROPE Framework Focuses on Mass Density

Expand ROPE Framework to Neutral Composition

1. Integration of multiple data sources supporting robust data assimilation and state estimation.[4]
2. Improved modeling of the drag coefficient due to connection with composition and temperature.
3. Step towards the inclusion ionosphere-thermosphere coupling within reduced order frameworks.
4. Opens possibility of incorporating additional physical constraints or use of the ROPE framework to drive physics-informed learning approaches.[5]

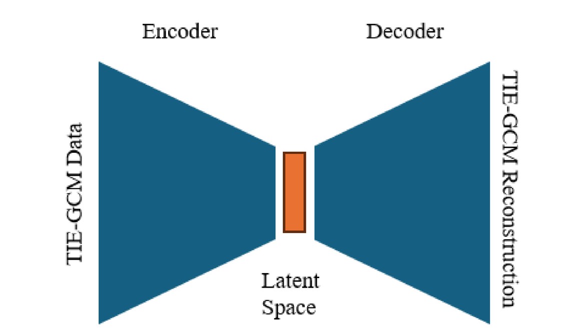
Methodology

ROPE → Dimensionality Reduction Focus

- TIE-GCM Neutral Species & Temperature (He, O_1, O_2, N_2, T_N)
  - Grid: [36, 72, 45] [LAT, LT, ALT] (100 km – 980 km) (Diffusive Equilibrium)
  - Total Dimensionality: 583,200

Convolutional Autoencoder

- Attention Mechanisms: No Attention (NA), Squeeze-Excitation (SE), Convolutional Block Attention Mechanisms (CBAM)
- Latent Space Loss Constraint
  - $\mathcal{L} = \mathcal{L}_{recon} + \alpha\,\mathcal{L}_{orth} + \beta\,\mathcal{L}_{corr}$
- Reconstruction Loss Scaling
  - Species Scaling
  - Altitude Loss Scaling
- Weighted Kp Input Sampler



Model Reconstruction MAPE Results

Table 1: Training, Validation, and Test Set Mean Absolute Percent Error using Combined Species PCA with 10, 20, and 30 Principal Components and Individual Species PCA with 10 Principal Components.

	Train				Validation				Test			
Components	10	20	30	10-I	10	20	30	10-I	10	20	30	10-I
He	8.39	6.05	4.96	5.83	10.24	7.45	6.07	6.64	9.25	6.71	5.61	6.40
O_1	6.67	5.00	4.26	6.41	7.69	5.80	5.05	6.51	6.61	4.97	4.33	6.16
O_2	19.51	14.50	11.82	18.16	22.35	16.20	13.33	18.12	19.26	14.23	11.83	17.48
N_2	16.22	12.30	10.20	14.43	17.55	13.14	11.04	14.22	16.95	12.81	10.87	15.01
T_N	1.25	0.82	0.67	0.94	1.54	0.96	0.79	1.04	1.44	0.88	0.73	0.99
Avg	10.41	7.73	6.39	9.15	11.87	8.71	7.26	9.30	10.70	7.92	6.68	9.21

Table 2: Comparison of Training, Validation, and Testing MAPE for CAE Models with 10 Latent Components, Loss Constraint = 0.005 and Attention.

	Train					Validation					Test				
Loss Constraint		α		β	α,β		α		β	α,β		α		β	α,β
Attention Mechanism	NA	SE	CBAM	NA	NA	NA	SE	CBAM	NA	NA	NA	SE	CBAM	NA	NA
He	4.25	7.40	5.78	3.89	3.89	9.22	10.09	9.91	7.93	9.42	4.94	7.51	6.26	4.56	4.64
O_1	3.78	6.23	5.24	3.50	3.43	4.95	7.29	6.18	4.66	4.75	4.18	6.55	5.72	3.83	4.01
O_2	9.16	14.51	11.69	8.62	8.49	12.39	18.83	17.48	12.89	12.74	10.24	14.69	12.68	10.00	10.06
N_2	7.41	11.99	9.77	6.89	6.79	9.84	15.03	13.35	9.87	10.08	8.20	12.54	11.07	7.98	8.02
T_N	0.71	1.08	0.88	0.65	0.65	0.91	1.34	1.08	0.88	0.90	0.83	1.23	1.01	0.77	0.80
Avg	5.06	8.24	6.67	4.71	4.65	7.46	10.52	9.60	7.25	7.58	5.68	8.50	7.35	5.43	5.50

A latent space of 10 was chosen in this work to provide better observability for downstream tasks such as data assimilation.

PCA MAPE Results

Average MAPE Decreases

- 10 → 20 : 25.98 %
- **10 → 30 : 37.57 %**
- 20 → 30 : 15.65 %
- 10 → 10-I : 13.92 %

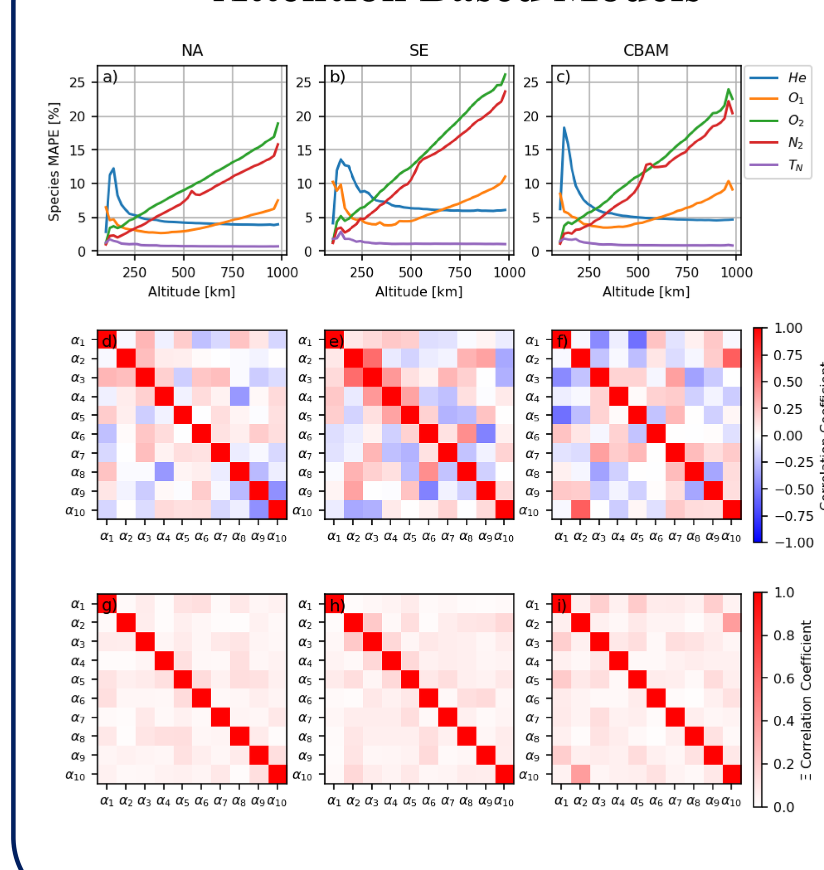
CAE MAPE Results

Average MAPE Decreases

- 10 PCA → α , NA : 46.98 %
- 10 PCA → α , SE : 20.51 %
- 10 PCA → α , CBAM : 31.31
- **10 PCA → β , NA : 49.25 %**
- 10 PCA → α, β, NA : 48.59 %
- **30 PCA → β, NA: 18.71 %**

Model Variation Results

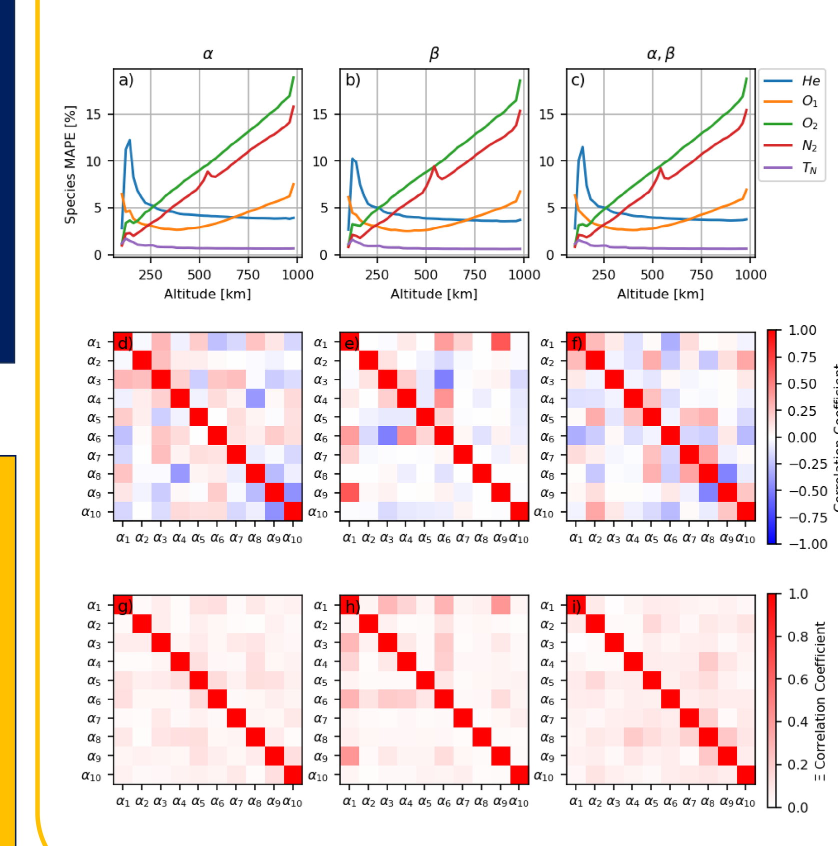
Attention Based Models



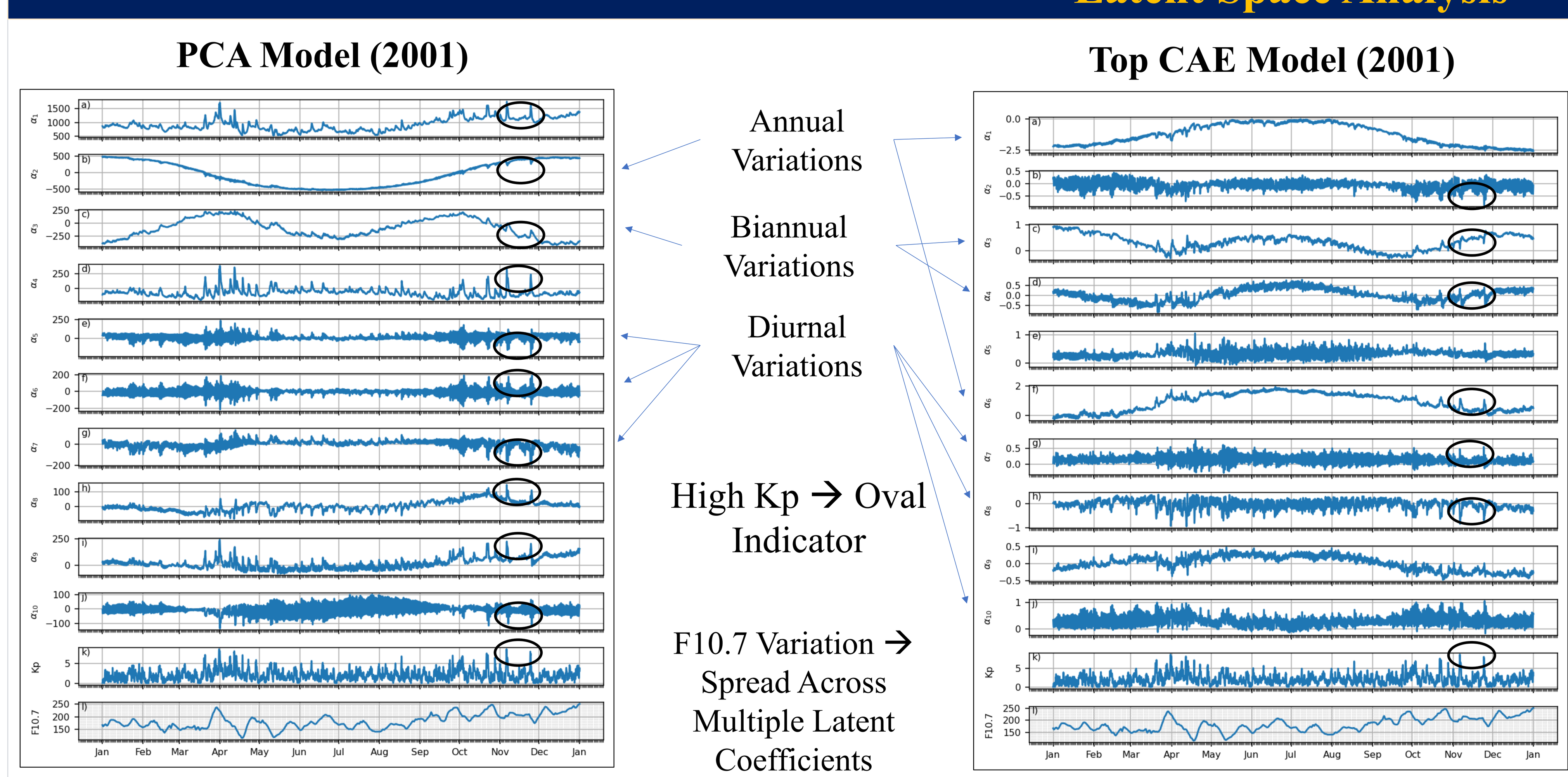
- Attention → Poorly Suited to Autoencoder Problem → Reconstruct Globally Distributed Physically Coupled System
- Attention Mechanisms → Slight Increase in Latent Space Correlations

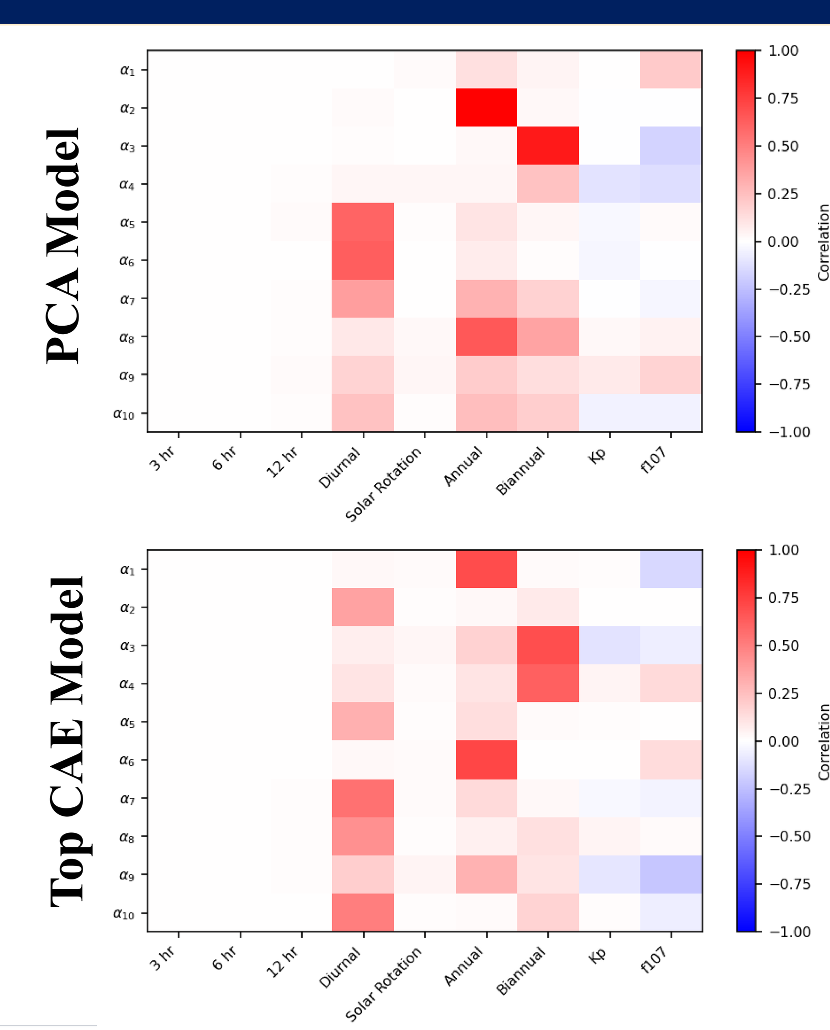
- All Models → Similar Reconstruction Error Dynamics
- All Models → Highly Uncorrelated Latent Space → Efficient Autoencoder
- Lowest latent space correlations in β

Latent Loss Scaling Based Models



Latent Space Analysis





Conclusions

- Dimensionality Reduction Model Compresses Neutral Species **583,200:10**
  - **5x Compression Ratio of TIE-GCM ROPE**
- Model with $\boldsymbol{\beta = 0.005}$ provides **best MAPE** and **Correlation Metrics**
- Best CAE Model outperforms equivalent PCA model by **49.25%** in Reconstruction
- Models without Attention outperform Models with Attention

- Similar Reconstruction Error Trends For All Models
  - **Reconstruction Trends → Physically Realistic**
  - **Heavy Species: ↑ MAPE ↑ Altitude → Rapid Number Density Decay → Small Denominators → Inflated Relative Error**
  - **Lighter Species: Lowest MAPE in Mid and High Altitudes**
  - **Low Reconstruction Error** for **Neutral Temperature Profile**

- Latent Space Highly Uncorrelated → Efficient Autoencoder
- Clear **Annual, Biannual, & Diurnal** Trends in Latent Space
- Limited to No trends **Sub-Diurnal** or **Solar Rotation** Trends in Latent Space

ACKNOWLEDGMENTS

The authors gratefully acknowledge support under NSF CAREER Award #2140204.

CONTACT

Nathaniel Michek, PhD, Research Scientist
ASSIST Lab, West Virginia University
Email nathaniel.michek@mail.wvu.edu

CONTACT

Piyush Mehta, PhD, Associate Professor
ASSIST Lab, West Virginia University
Email piyush.mehta@mail.wvu.edu

References

1) Mehta et al. *TIE-GCM ROPE Part I, DOI: 10.1029/2024SW004185*
2) Tapedia et al., *TIE-GCM ROPE Part II,* DOI: *10.1029/2025SW004661*
3) Sicoli et al., *ROPE: A Reduced-Order Probabilistic Emulator for Fast and Uncertainty-Aware Thermospheric Density Emulation,* Poster #8, Apr. 29, 2026
4) Narayanan et al, *Bridging High-Fidelity Modeling and Operations with Reduced-Order Thermospheric Data Assimilation,* Poster #3, Apr. 29, 2026
5) Ulifan et al., *Physics-Informed ML Framework for Thermosphere Density Prediction,* Poster #6 , Apr. 29, 2026