Exploring Adversarial Storylines of Tropical Hurricanes

Marco **Froelich**[1,2], Sebastian Engelke[2]

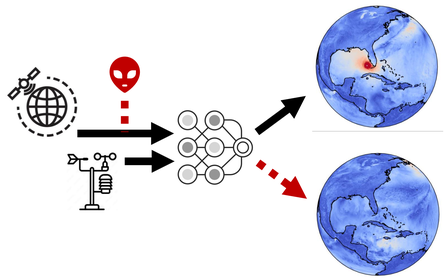
[1]marco.froelich@unige.ch, [2]University of Geneva, Research Institute for Statistics and Information Science

Motivation

AI models revolutionized weather forecasting by improving numerical weather prediction skill and with computational costs orders of magnitude smaller. In addition, backpropagation allows direct computation of model sensitivities to initial conditions. Adversarial attacks leverage these sensitivities to influence the output of the prediction system by finding optimal initial condition perturbations.

The **two perspectives** of adversarial attacks in AI weather forecasting:

- Operational systems are susceptible to malicious attacks [1 - 4]
- Adversarial attacks are efficient tools for:
  - Studying atmospheric predictability [5, 6]
  - Exploration of **extreme events** [7, 8]



Methods

We investigate adversarial attacks of **GraphCast-AMSE** [9, 10] on deviating tropical hurricane tracks by *minimizing mean sea-level pressure* at diverse target locations. An L^{∞} penalty over each normalized variable and pressure-level field – approximated by Softmax weights – is introduced to remain close to HRES-fc0 initialization. GraphCast-AMSE takes two input states, but we limit perturbations to the second state only.

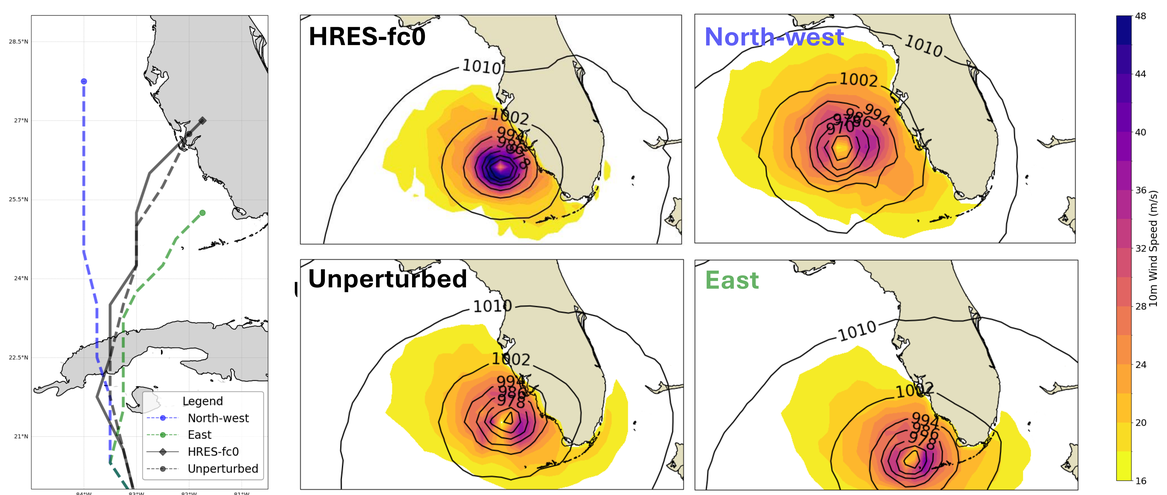
Objective:

$$\mathcal{L}(\boldsymbol{X}) = MSLP_{\theta_T,\phi_T,t} + \frac{1}{|\mathcal{V}|}\sum_{v\epsilon\mathcal{V}}\sum_{\theta,\phi}\omega_{\theta,\phi}(\delta^{(v)}) \cdot |\delta^{(v)}_{\theta,\phi}|$$

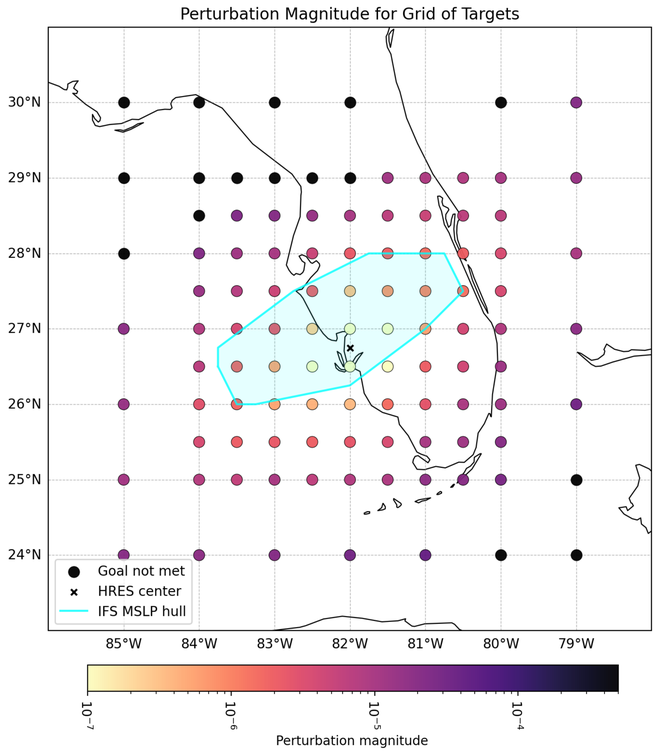
where $\omega_{\theta,\phi}$ are Softmax weights at θ latitude and ϕ longitude (0.25°) computed over the spatial grid; t the final timestep; $\mathcal{V}$ the set of diagnostic and prognostic input variables; δ the perturbation.

Successful Attacks

Forecast initialized on 2022-09-25T18:00 | Valid on 2022-09-28T12:00 | mslp max perturbation ~ 0.4 hPa

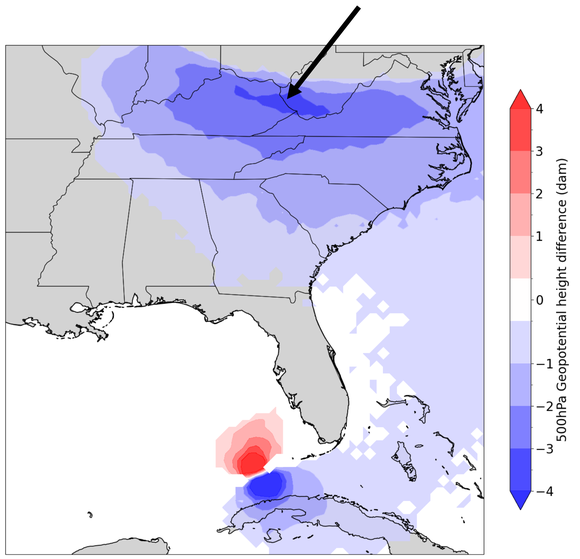


Occam's Razor uncertainty



Mesoscale drivers

[1] attribute NWP track forecast North-western bias to an underestimation of a **shortwave trough**, which Eastern adversarial storylines correct.



Conclusions

- **Diverse storylines** of tropical hurricanes generated at 0.25˚resolution
- **Hurricane Ian**: driving mesoscale features of eastern-moving tracks recovered → growing confidence in AI model outputs
- Initial results in **track uncertainty forecasting** based on perturbation magnitudes

? Are perturbations meaningful? Do they lead to correct track deviations in **other models**?
? What uncertainty is being addressed?

References

[1] Imgrund, E., Eisenhofer, T., Rieck, K., 2025. Adversarial Observations in Weather Forecasting.
[2] Deng, Y., Santos, F., Tan, P.-N., Luo, L., 2025. Adversarial Attacks on Downstream Weather Forecasting Models: Application to Tropical Cyclone Trajectory Prediction.
[3] Arif, H., Chen, P.-Y., Gittens, A., Diffenderfer, J., Kailkhura, B., 2025. Forecasting Fails: Unveiling Evasion Attacks in Weather Prediction Models.
[4] Chichifoi, K., Merizzi, F., Colajanni, M., 2025. Evaluating Adversarial Attacks on Federated Learning for Temperature Forecasting. https://doi.org/10.48550/arXiv.2512.13207
[5] Vonich, P.T., Hakim, G.J., 2024. Predictability Limit of the 2021 Pacific Northwest Heatwave From Deep-Learning Sensitivity Analysis. Geophysical Research Letters 51, e2024GL110651.
[6] Vonich, P.T., Hakim, G.J., 2025. Testing the Limit of Atmospheric Predictability with a Machine Learning Weather Model.
[7] Whittaker, T., Luca, A.D., 2026. Constructing Extreme Heatwave Storylines with Differentiable Climate Models. Weather Clim. Dynam. 7, 393–410.
[8] Hakim, G.J., Agrawal, A., 2026. Gray Swan Factory: Making Extreme Events from Ordinary Cyclones.
[9] Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W., Merose, A., Hoyer, S., Holland, G., Vinyals, O., Stott, J., Pritzel, A., Mohamed, S., Battaglia, P., 2023. Learning skillful medium-range global weather forecasting. Science 382, 1416–1421.
[10] Subich, C., Husain, S.Z., Separovic, L., Yang, J., 2025. Fixing the Double Penalty in Data-Driven Weather Forecasting Through a Modified Spherical Harmonic Loss Function.
[11] Colby, F.P., Barlow, M., Penny, A.B., 2024. Steering Flow Sensitivity in Forecast Models for Hurricane Ian (2022). Weather and Forecasting 39, 821–829.