Probing The Physical Limits of Extreme Events with Differentiable Weather Models

Gregory J. Hakim

University of Washington

Workshop on Unpacking AI Weather Emulators
1 October 2026

Collaborators: Aishwarya Agrawal (UC Berkeley) and Ya-Shin Chi (U Washington)

What **new science** can we do now that we could not before?

- predictability (Pedram, Karen, & Aditi's talks)
- wholistic system dynamics (Theo's talk on mechanistic interp)
- novel data assimilation strategies
- hypothesis testing using deep nonlinear optimization

Deep Nonlinear Optimization

- Sensitivity review
  - OLD way: adjoints and shallow searches on quadratic functionals
  - NEW way: autograd and deep nonlinear searches

- Gray swans, and the physical limit of extreme events

Traditional Gradient Descent

Outer loop:

(1) nonlinear forward integration

(2) construct the tangent-linear model about this solution

Inner loop:

(1) forward integration with the TLM, compute J

(2) backward integration with the adjoint model, compute grad(J)

(3) increment the state with the gradient to reduce J

(4) return to (1)

$\Longrightarrow$ periodically return to the **outer loop** and re-linearize (expensive)

Pro: inner loop has a unique minimum (quadratic functional)

Con: TLM & adjoints codes (nasty); expensive— few outer loops (shallow searches)

Autograd Descent

```
@partial(jax.checkpoint, static_argnums=(3,))
def forecast(model,f_encoded,forcings,n):
    # advance state to the next verification time by marching one internal timestep at a time
    def scan_fn(f_encoded,el):
        f_encoded = model.advance(f_encoded,forcings)
        return f_encoded,f_encoded
    f_encoded, all_encoded = lax.scan(scan_fn,f_encoded,jnp.arange(n))
    return f_encoded
```

Single loop:

(1) nonlinear forward pass, compute J
(2) backpropagate grad(J)
(3) increment the state with the gradient to reduce J
(4) return to (1)

Pro: very deep searches
Con: non-unique minima

Gray Swan Factory

Gray swan:

- extreme event that is plausible, but thought to be unlikely
- cf. black swan: probability so low that the event is unknown

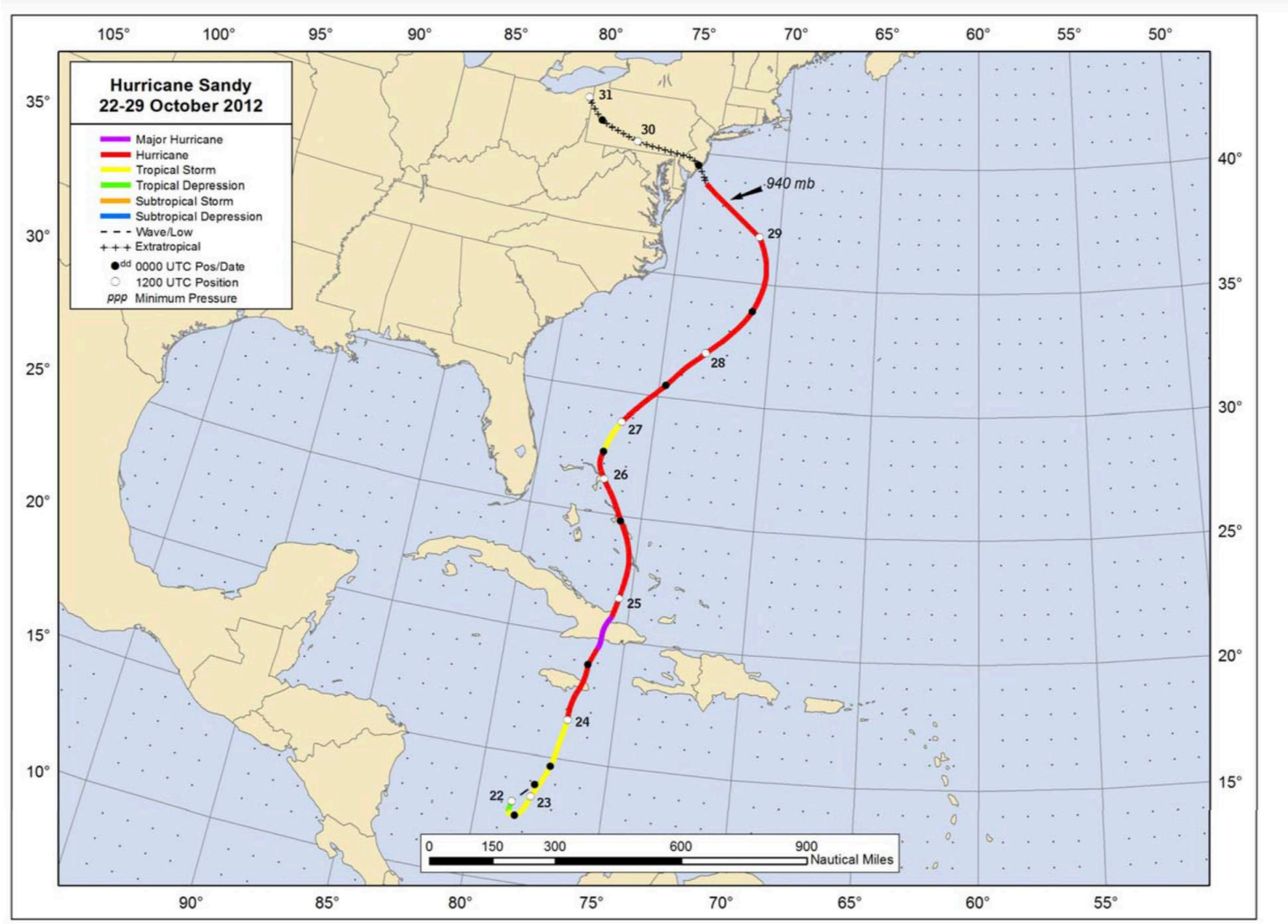
Hypothesis:

- these events can be manufactured using directed gradient descent
- synthetic extremes for ML model training/testing outside of ERA5
- explore the physical limits of extreme events (~impossible with ensembles)

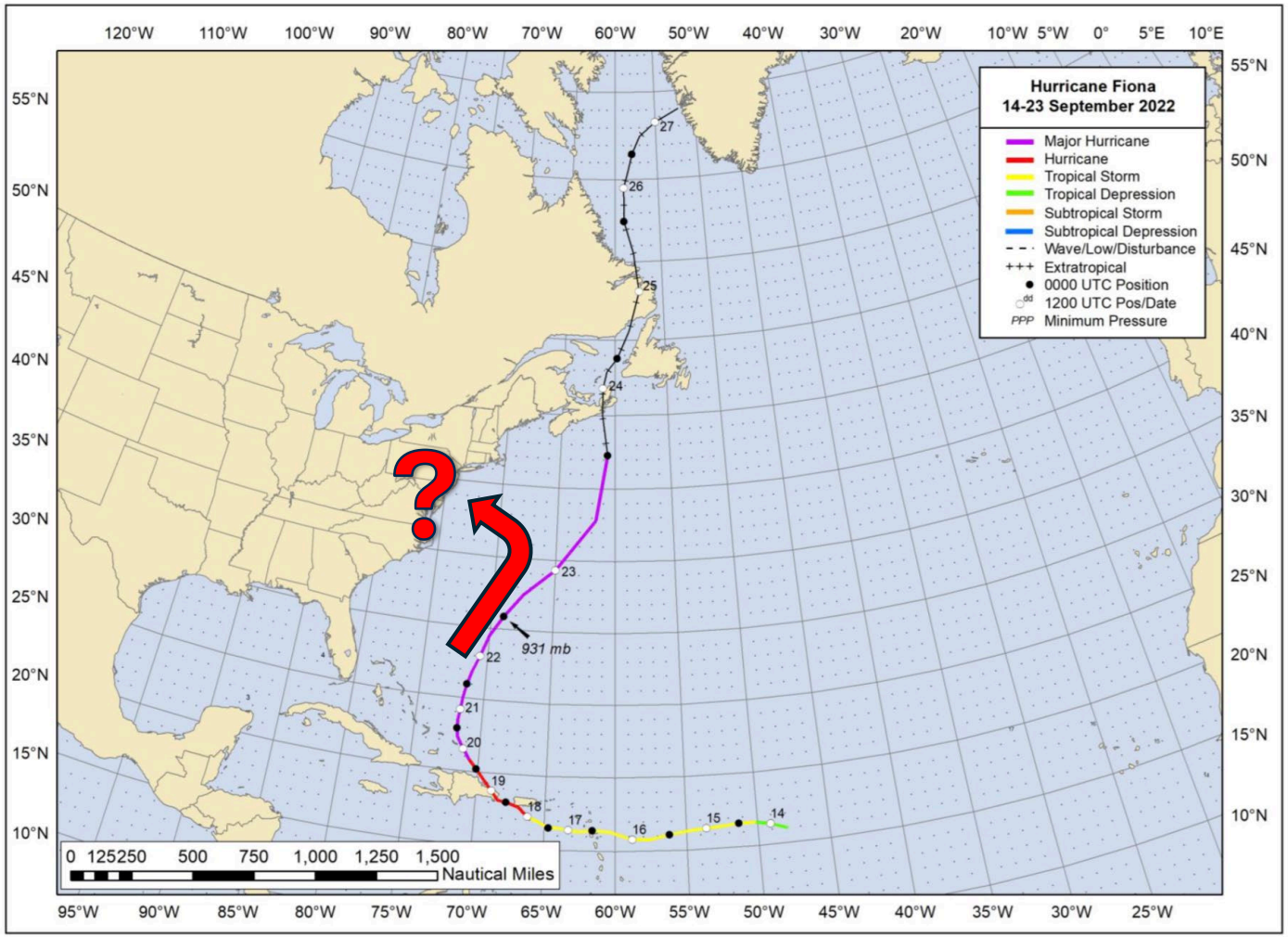
Experiment:

- "ordinary" Atlantic hurricanes have nearby analogs similar to hurricane Sandy (2012)
  - **make small changes to ERA5** to achieve a gray swan outcome
  - $10^{30} \rightarrow 1$!
- method: nonlinear gradient descent using NeuralGCM at 2.8° resolution

Hurricane Sandy (2012)



Hurricane Fiona (2022)

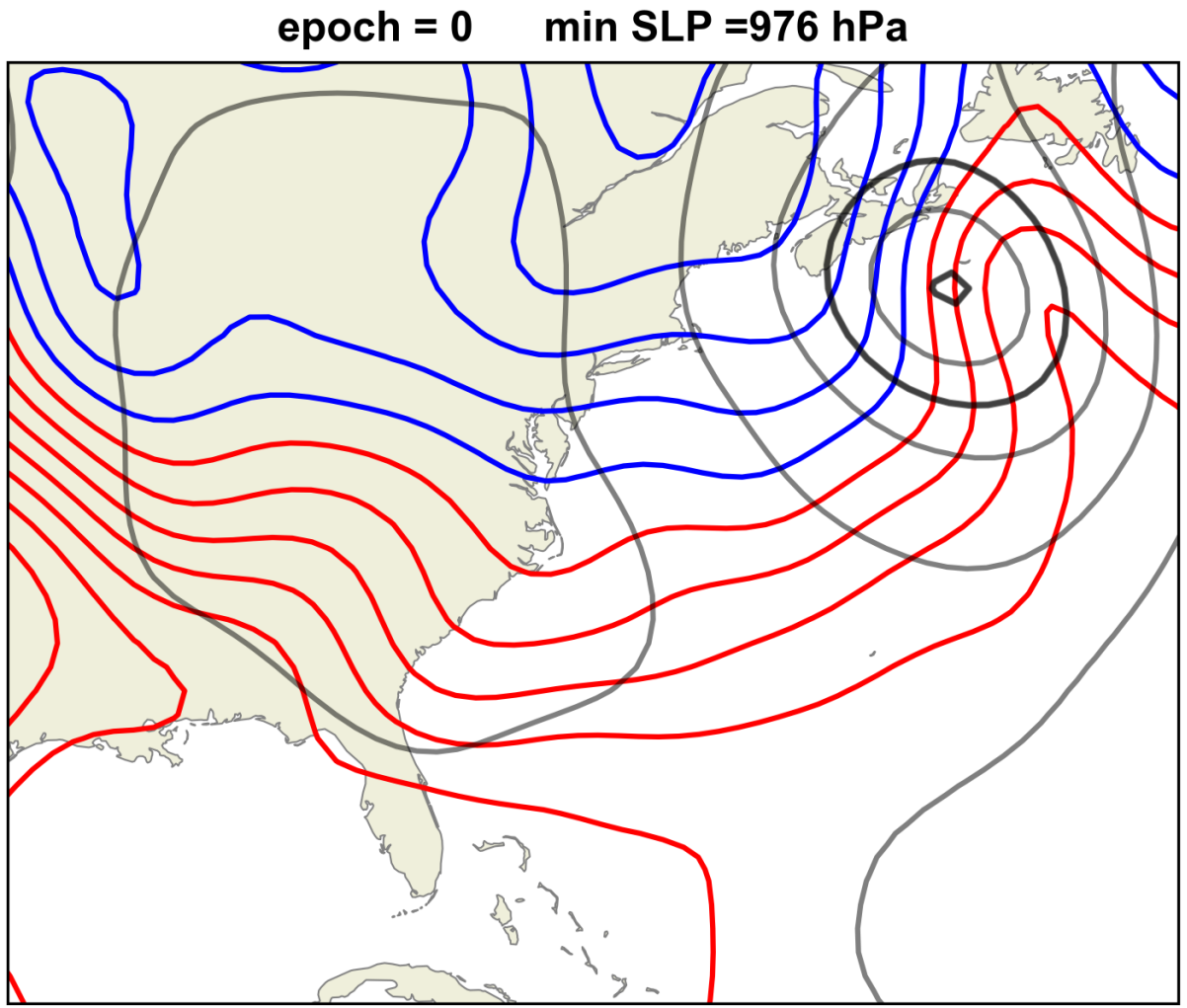


Control Initial Conditions from ERA5

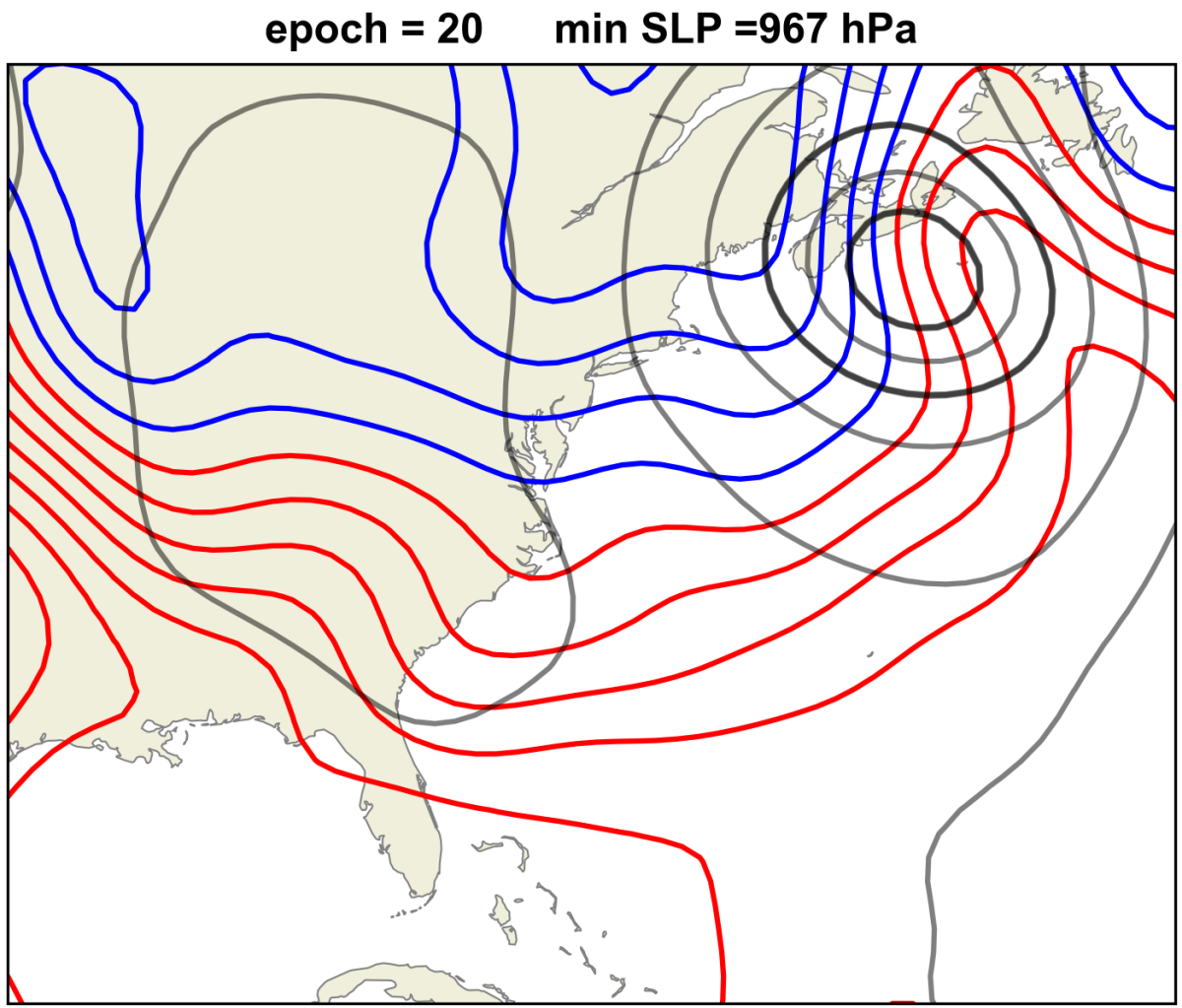
Control Forecast from ERA5

Control Forecast from ERA5

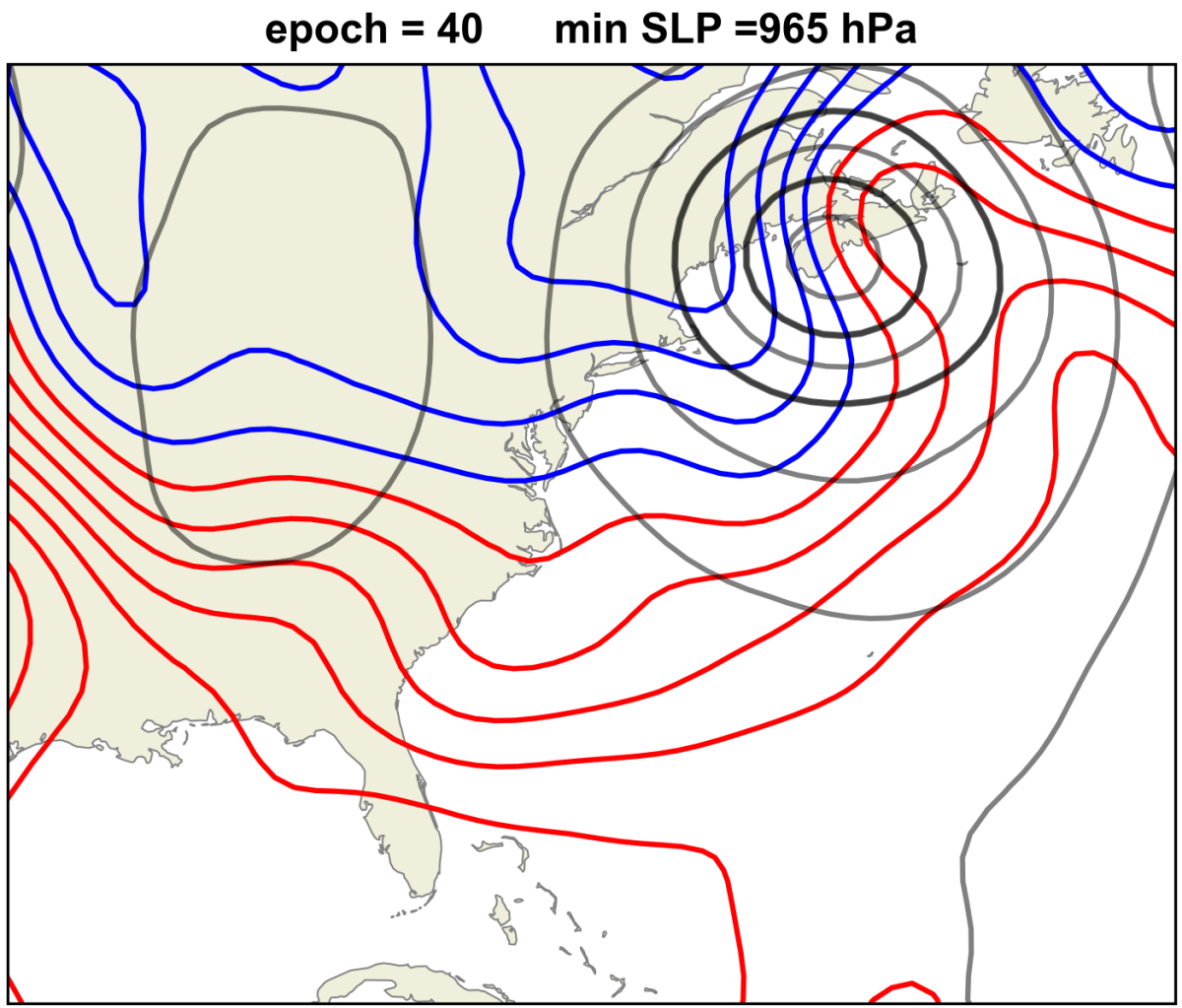
MSLP every 8 hPa
- **bold** every 16hPa
1000 hPa T every 2°C
- red > 19°C



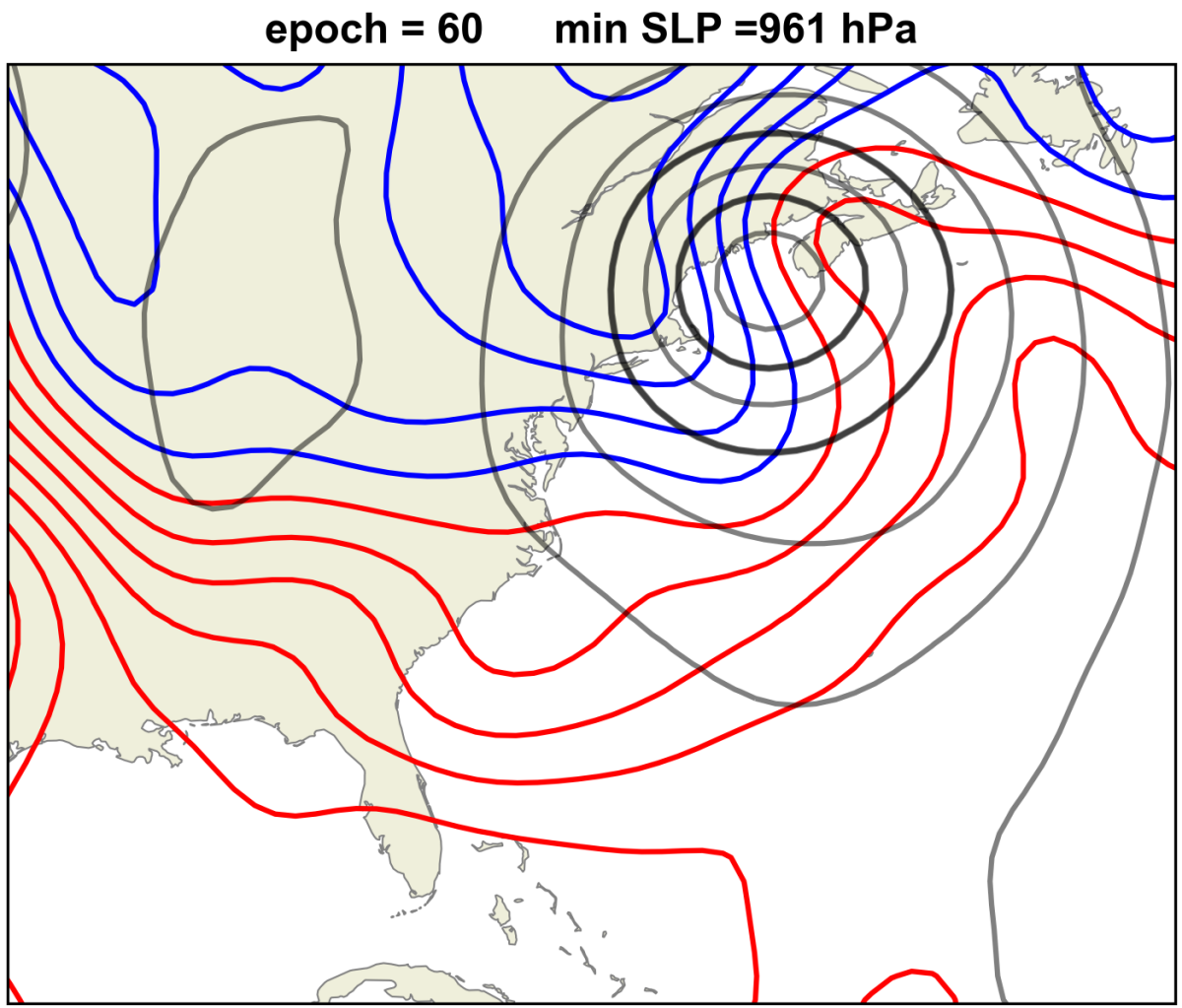
Gray Swan Forecast



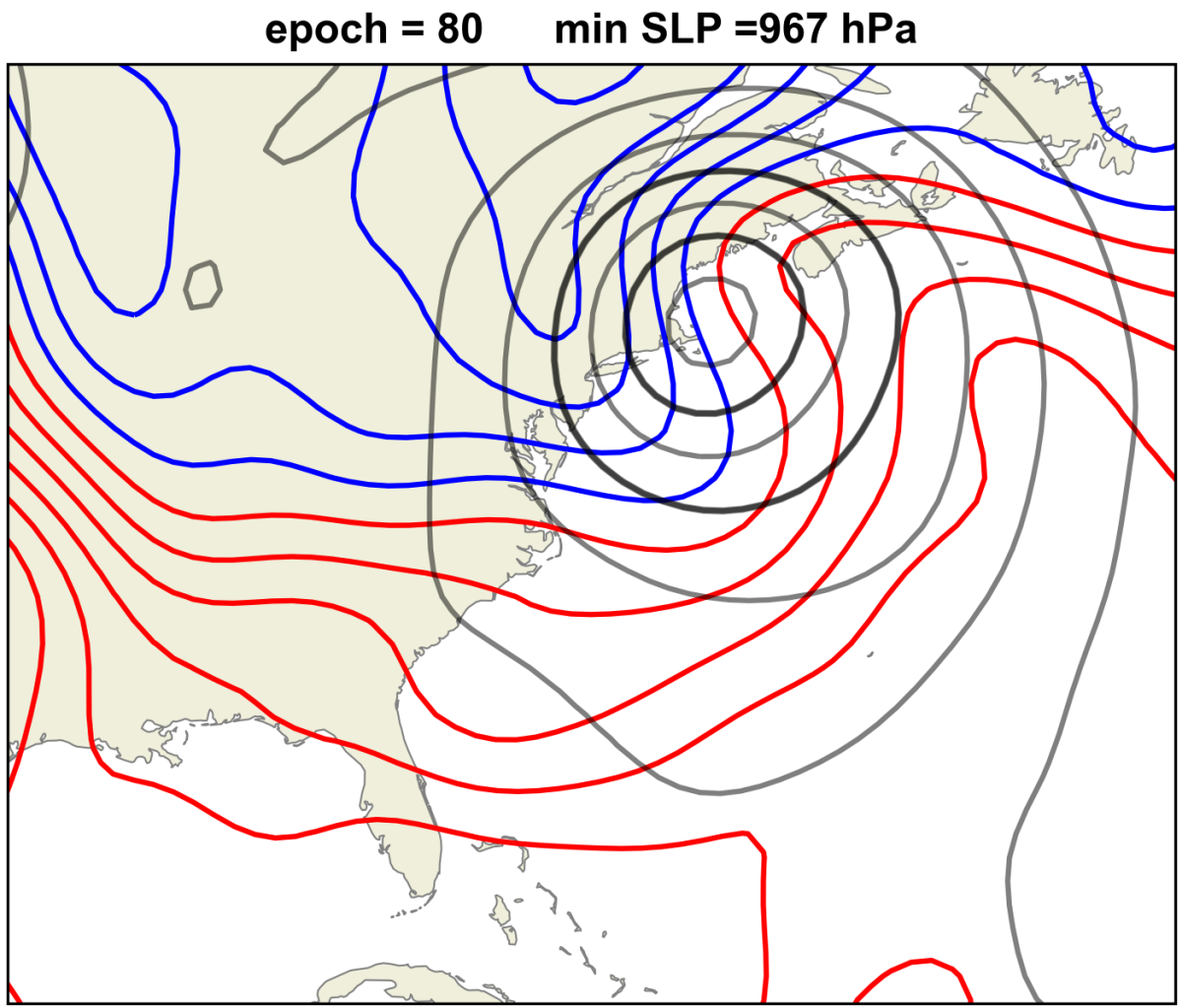
Gray Swan Forecast



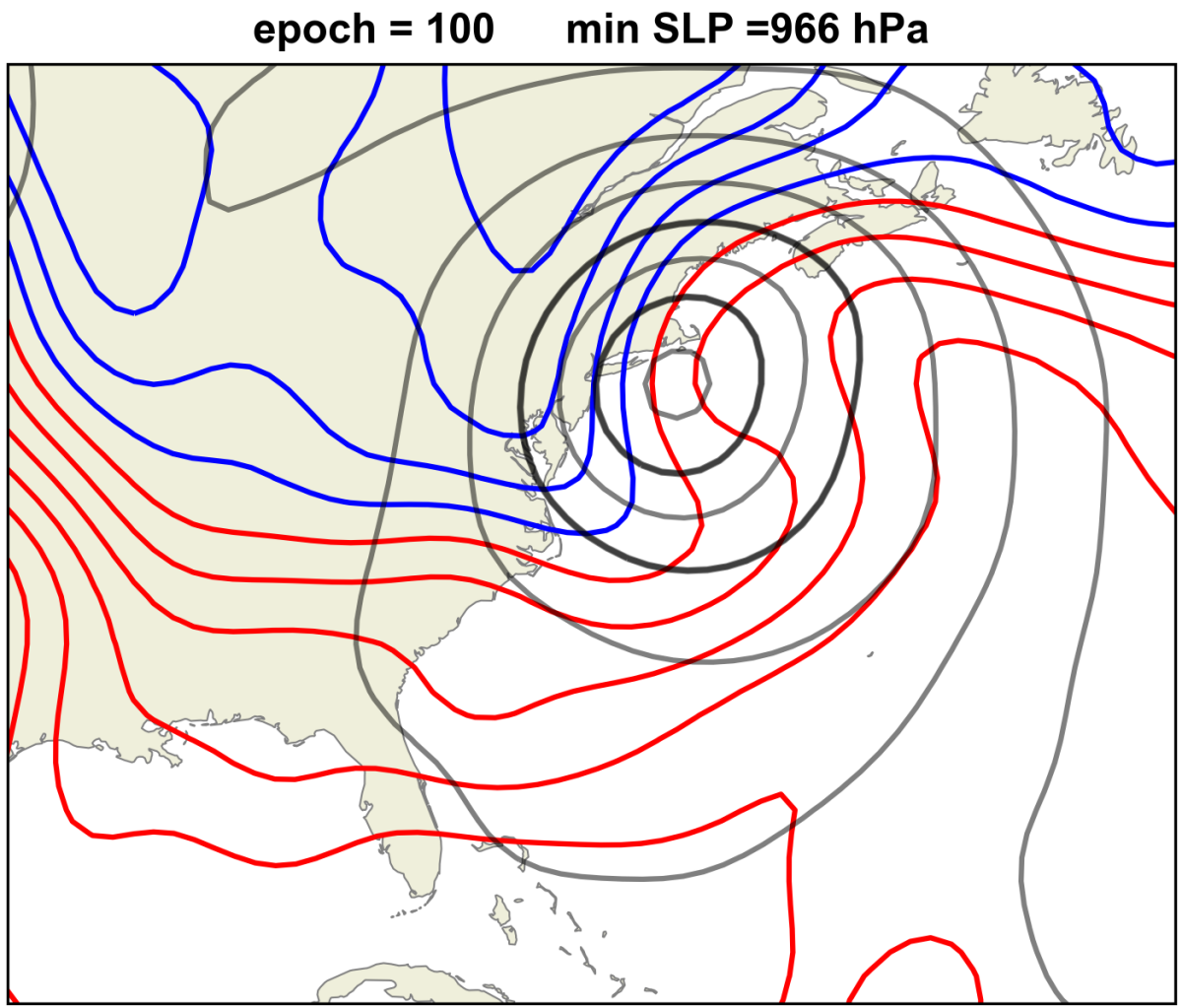
Gray Swan Forecast



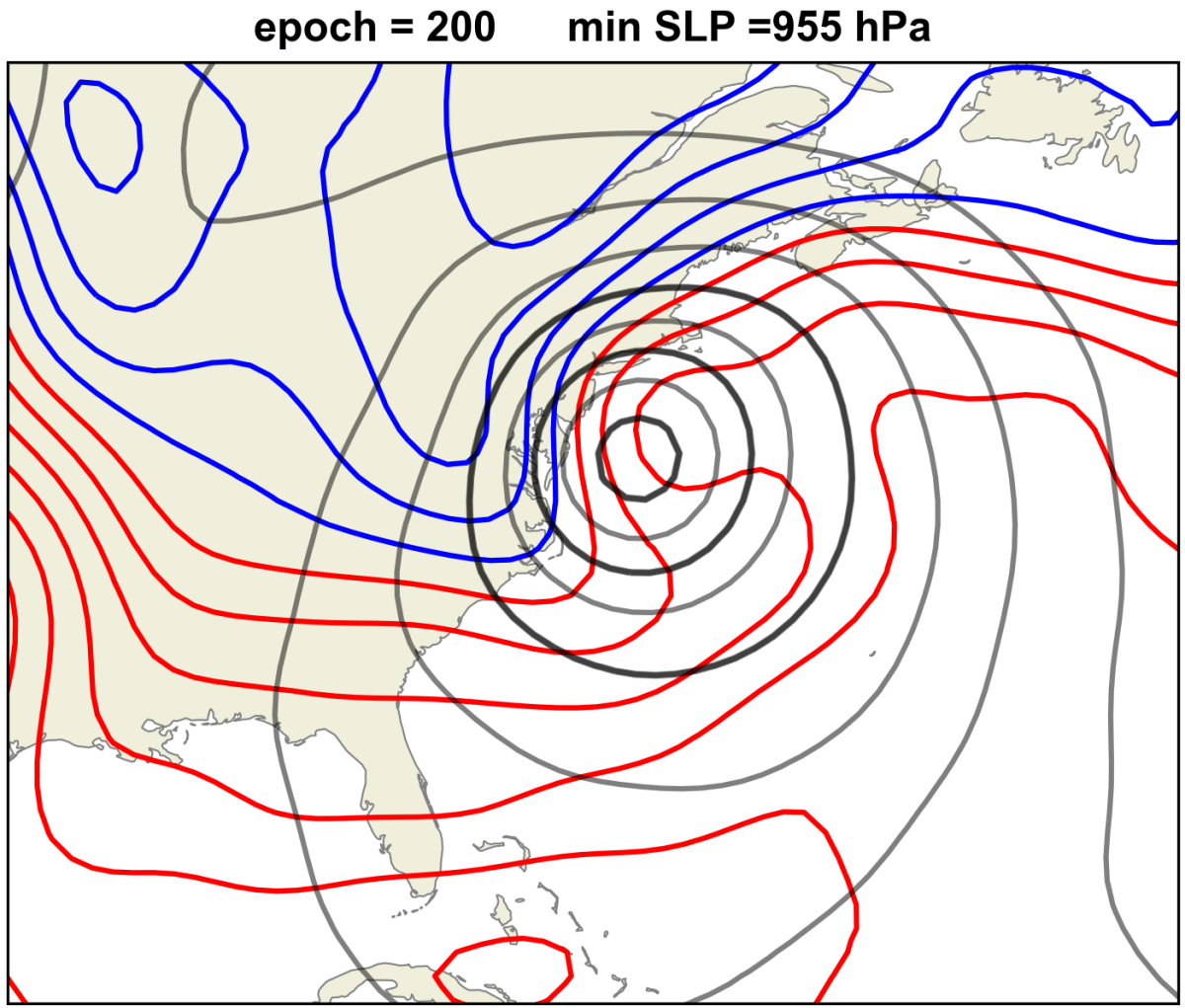
Gray Swan Forecast



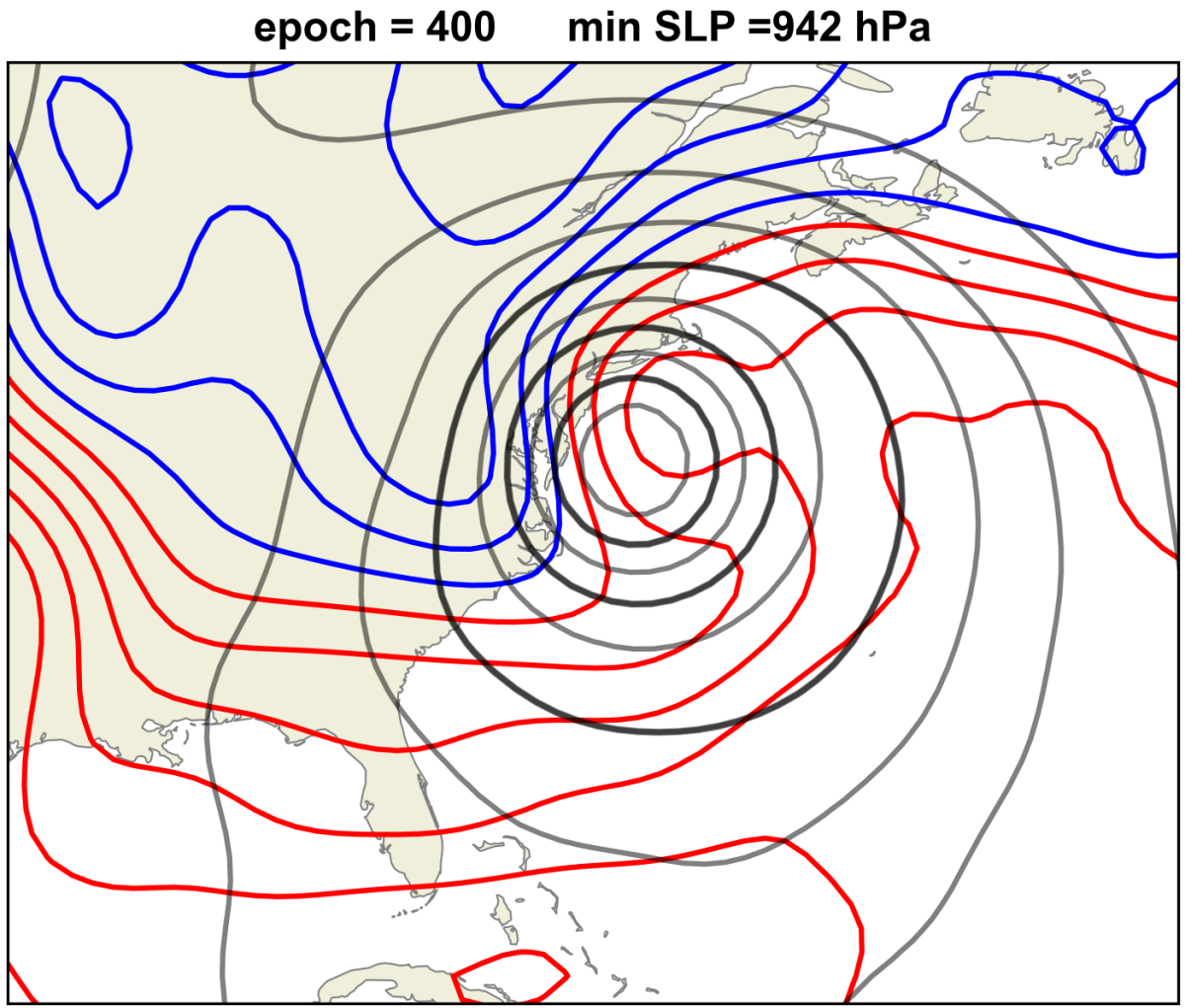
Gray Swan Forecast



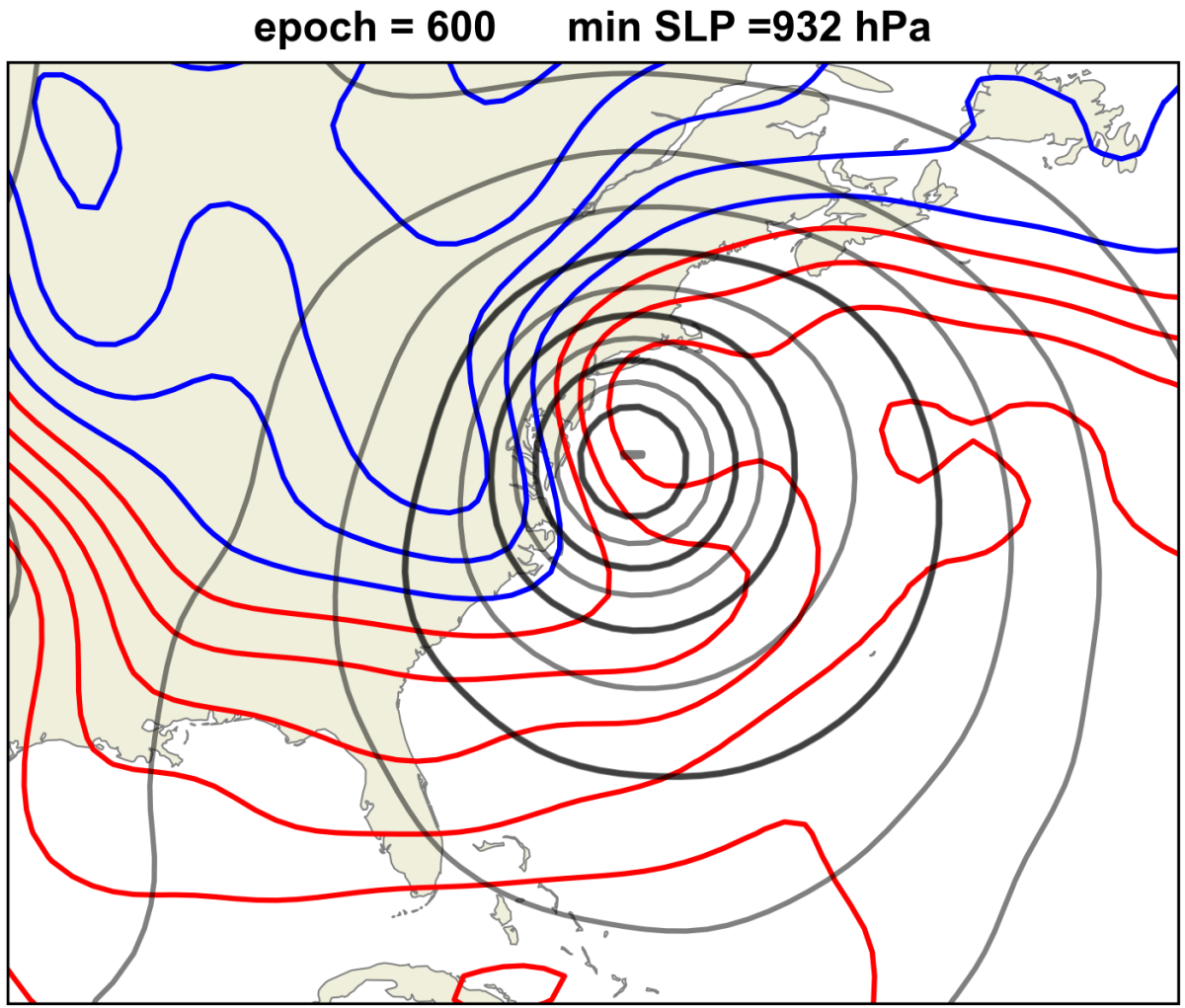
Gray Swan Forecast



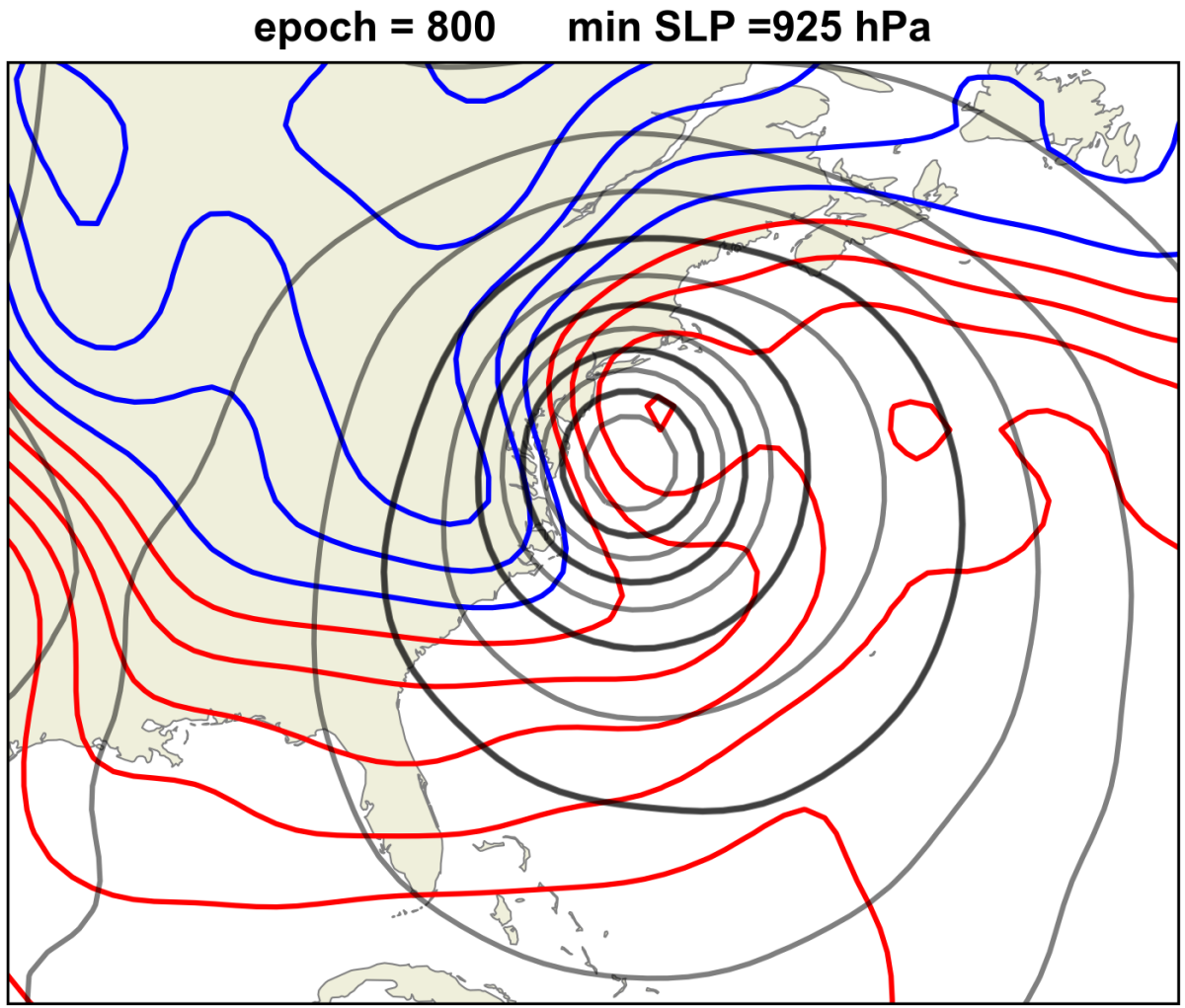
Gray Swan Forecast



Gray Swan Forecast



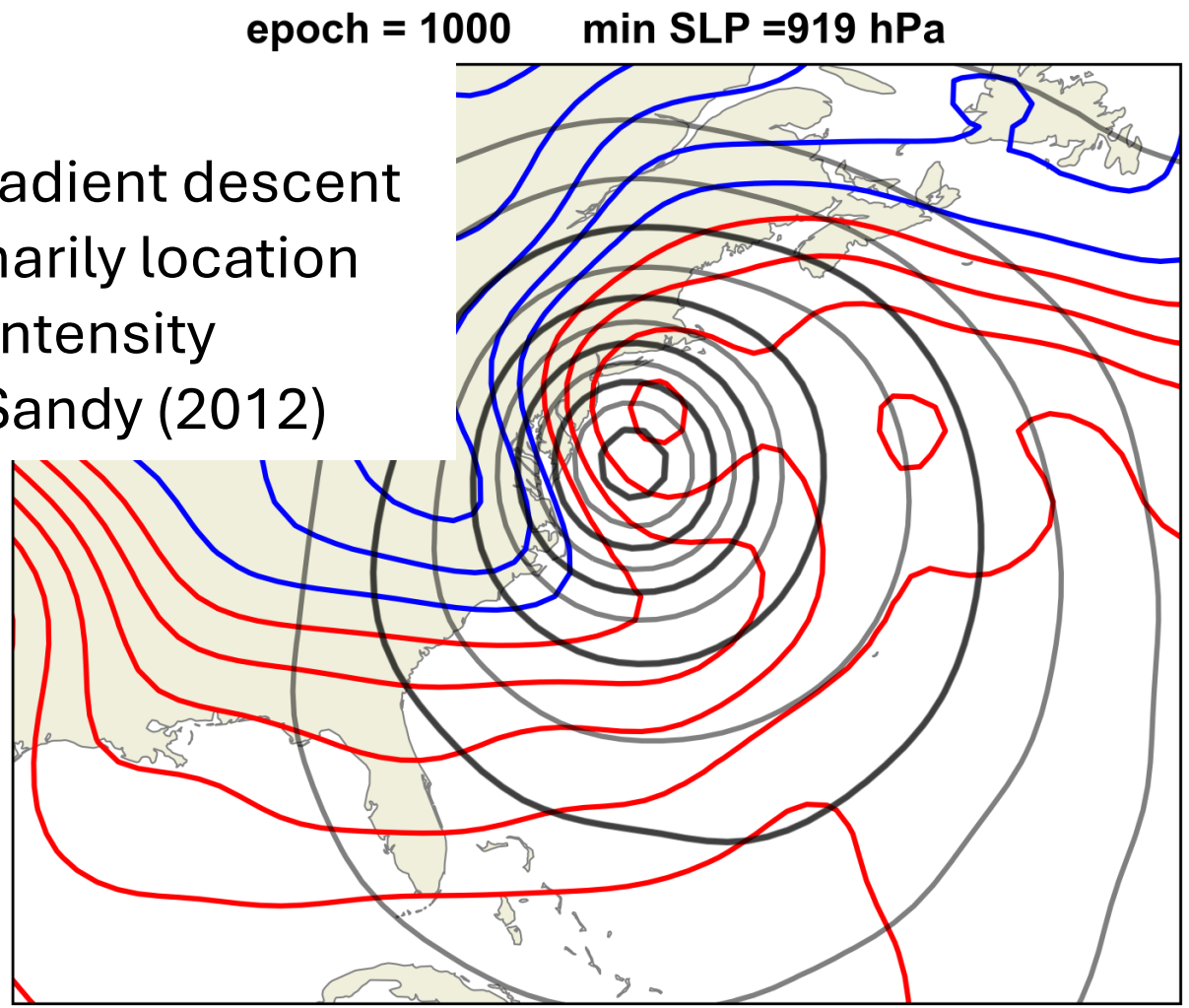
Gray Swan Forecast



Gray Swan Forecast

Gray swan solution:

- smooth function of gradient descent
- epochs 50—100: primarily location
- epochs>100: mostly intensity
- ~20hPa deeper than Sandy (2012)



control forecast

gray swan forecast

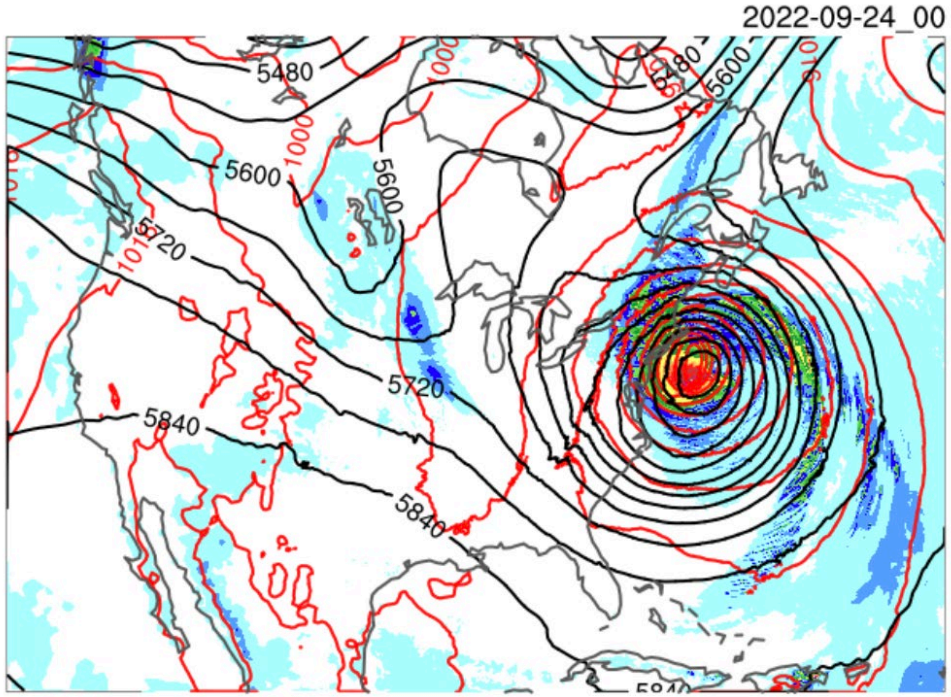
t = 48h

Initial Conditions & Partitioned Solution

t = 0h

t = 48h

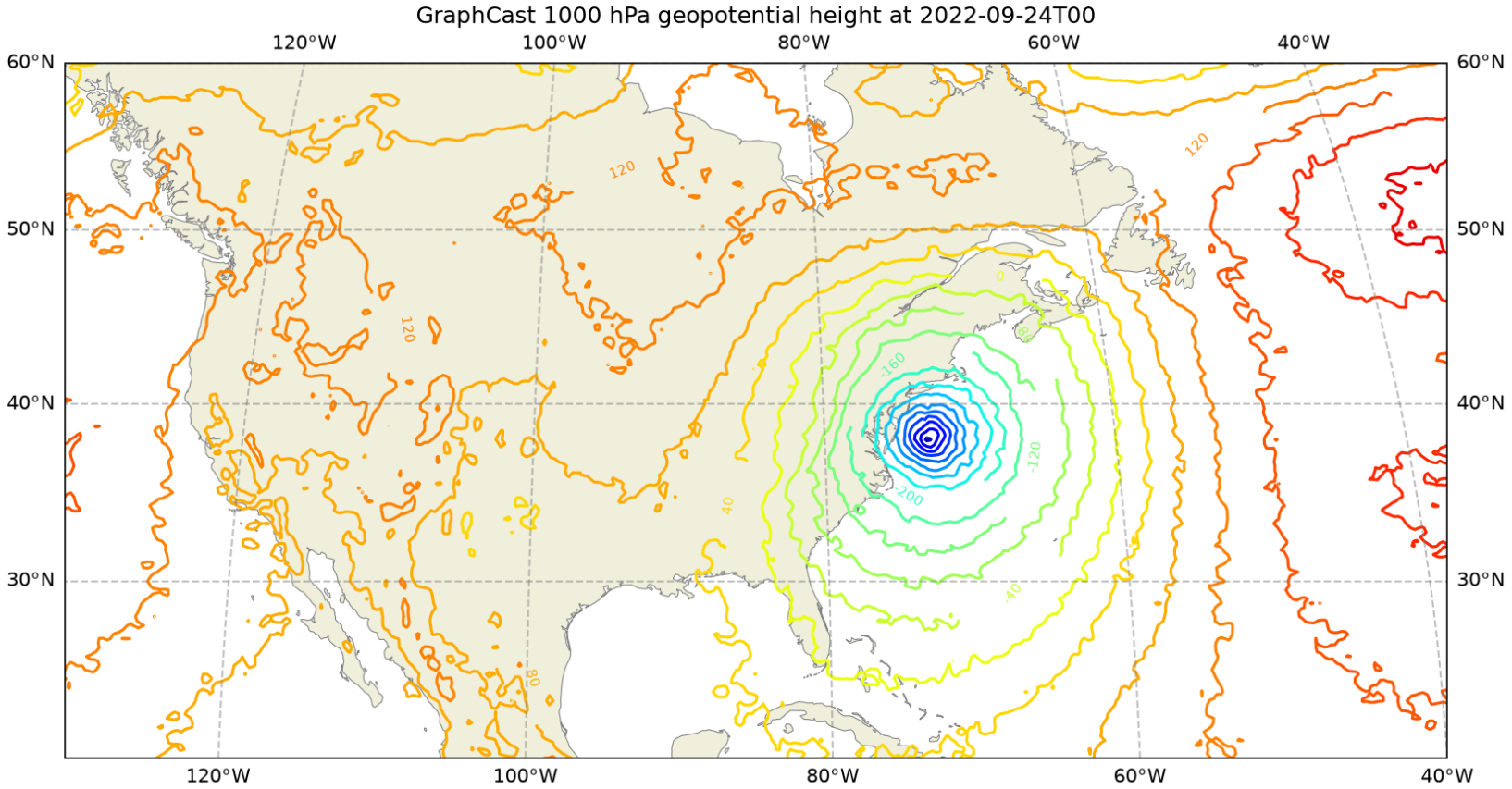
NGCM Gray Swan integrated in **MPAS**





Ya-Shin Chi

Graphcast Gray Swan (pure ML)



What has it learned? It can produce a storm stronger than any observed

Non-unique solutions

initial intensity

dots every 10^3 epochs

final intensity

ERA5

$$J = (x_G - g)^2 + \frac{\alpha_i}{N} \sum (\mathbf{x}_T - \overline{\mathbf{x}}_T)^2$$

Summary & Outlook

- **Autograd gradient descent**
  - fully nonlinear hypothesis testing
  - deep, computationally efficient, searches
- **Atmosphere's attractor is dense with Gray swan extreme events**
  - these near-miss storms having intensity greater than previously observed
  - a pure ML model (Graphcast) does this too
  - useful for "what if" testing; synthetic training data
  - explore the physics of these limiting events