

Assessing the Skill and Sensitivity of AI-Generated Global NWP Emulators in the NOAA HWT Spring Forecasting Experiment

David Harrison^{1,2}, Tim Supinie², Israel Jirak², Adam Clark³

1. Cooperative Institute for Severe and High-Impact Weather Research and Operations, University of Oklahoma, Norman, OK
2. NOAA/NWS/Storm Prediction Center, Norman, OK
3. NOAA/OAR/National Severe Storms Laboratory, Norman, OK



1. Motivation:

AI-based numerical weather prediction (NWP) emulators have been shown to exceed the skill of traditional NWP models while requiring fewer computational resources and time to run. However, these systems have only been evaluated in a limited capacity within an operational setting. How do these emulators compare to NWP in an operational, real-time environment?

- Do the emulators provide utility for the prediction of severe weather at extended lead times?
- How do different initial conditions impact emulator performance?

2. Experiment Design:

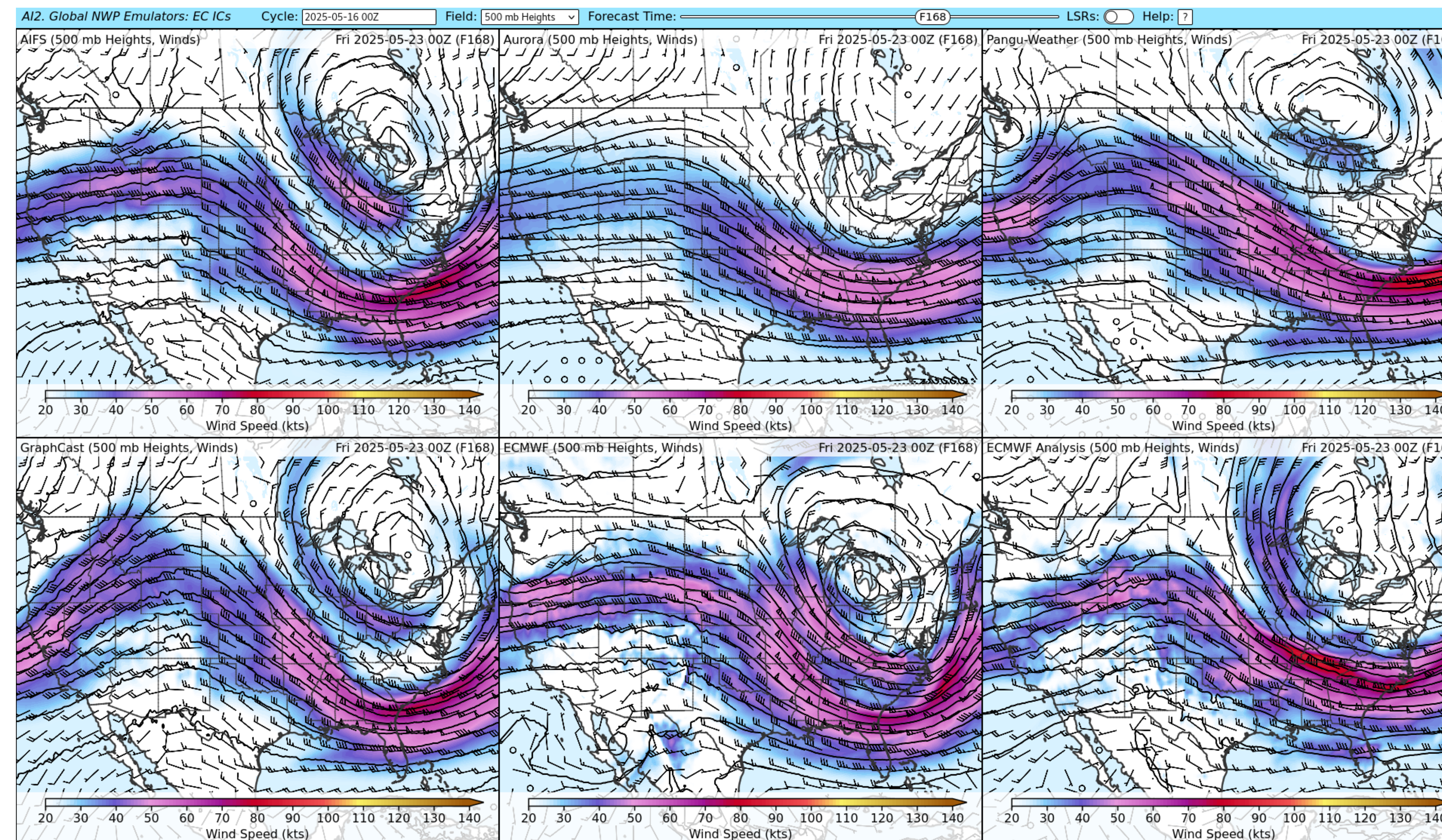
Participants of the 2025 Hazardous Weather Testbed (HWT) Spring Forecasting Experiment (SFE) evaluated the skill and utility of five deterministic AI emulator forecasts at 7-day ((F156 – F180) lead times. These emulators were initialized with either GFS or EC initial conditions (ICs) and compared to the GFS or ECMWF analysis respectively. The operational GFS and ECMWF were also included in the evaluations. Model names were hidden on the interactive webpage so that participants did not know which forecast was created by which model during evaluations.

GFS ICs:

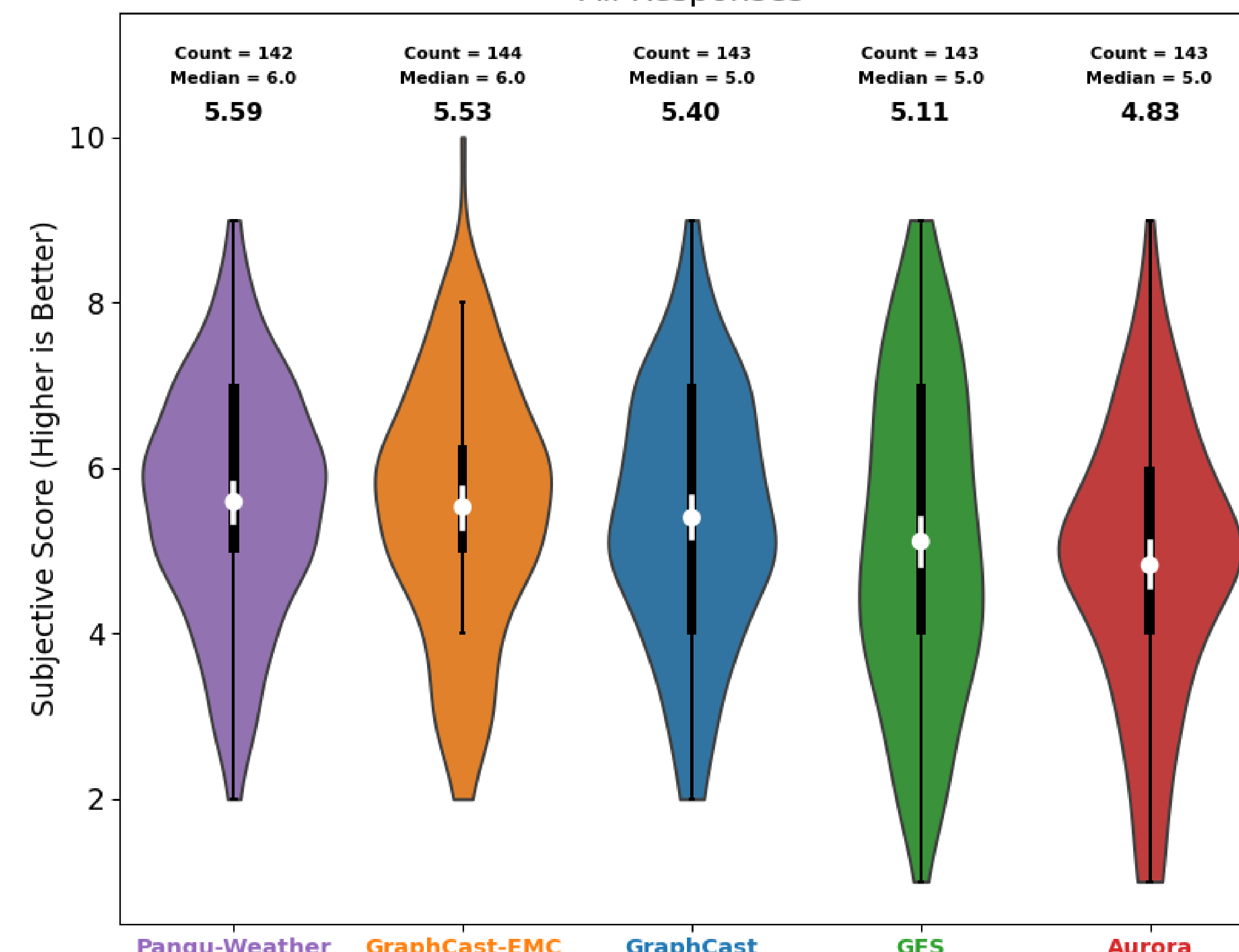
- GraphCast
- GraphCast-EMC
- Pangu-Weather
- Aura

EC ICs:

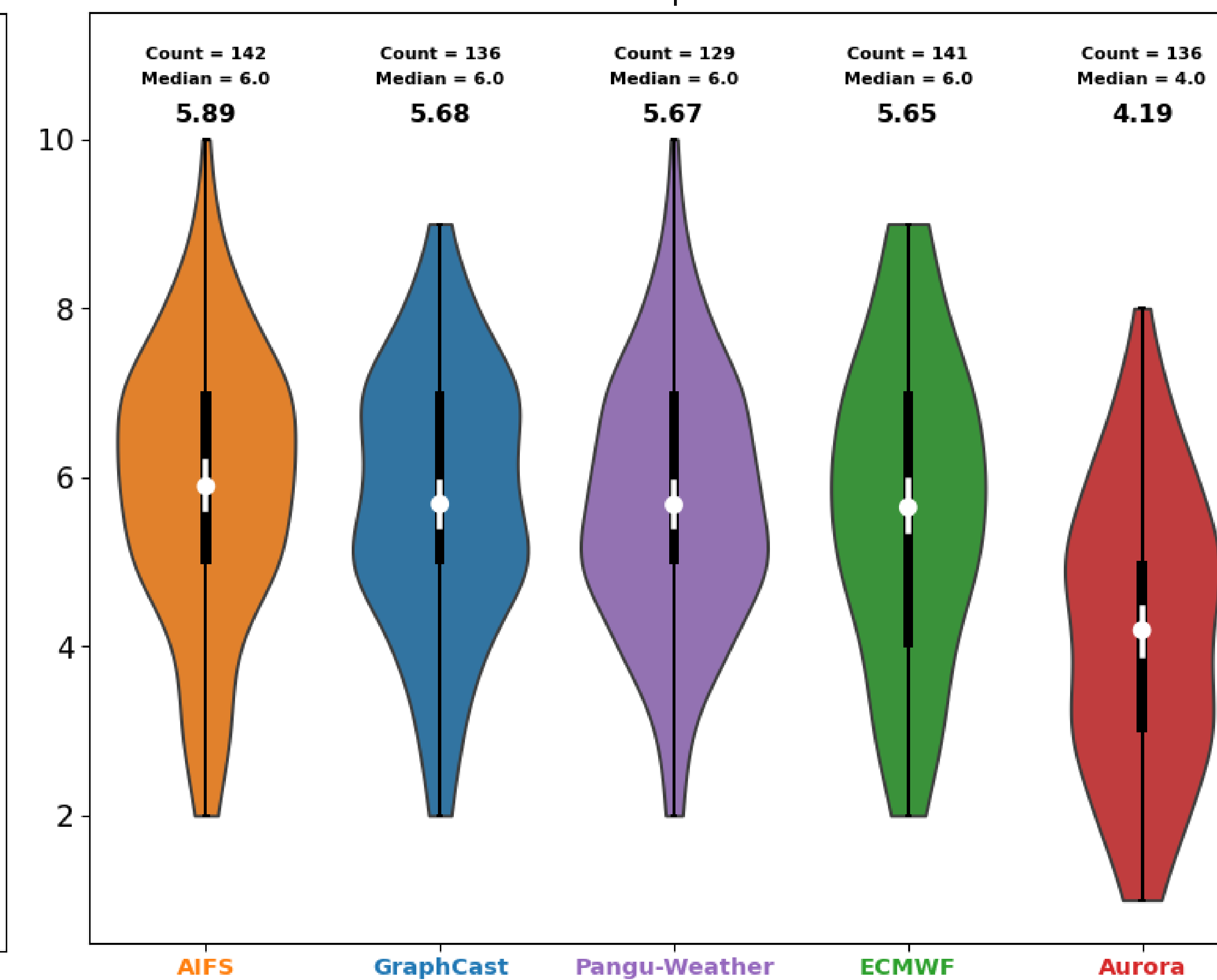
- GraphCast
- AIFS
- Pangu-Weather
- Aura



Global NWP Emulators with GFS ICs: Day 7 Synoptic Pattern
All Responses



Global NWP Emulators with EC ICs: Day 7 Synoptic Pattern
All Responses

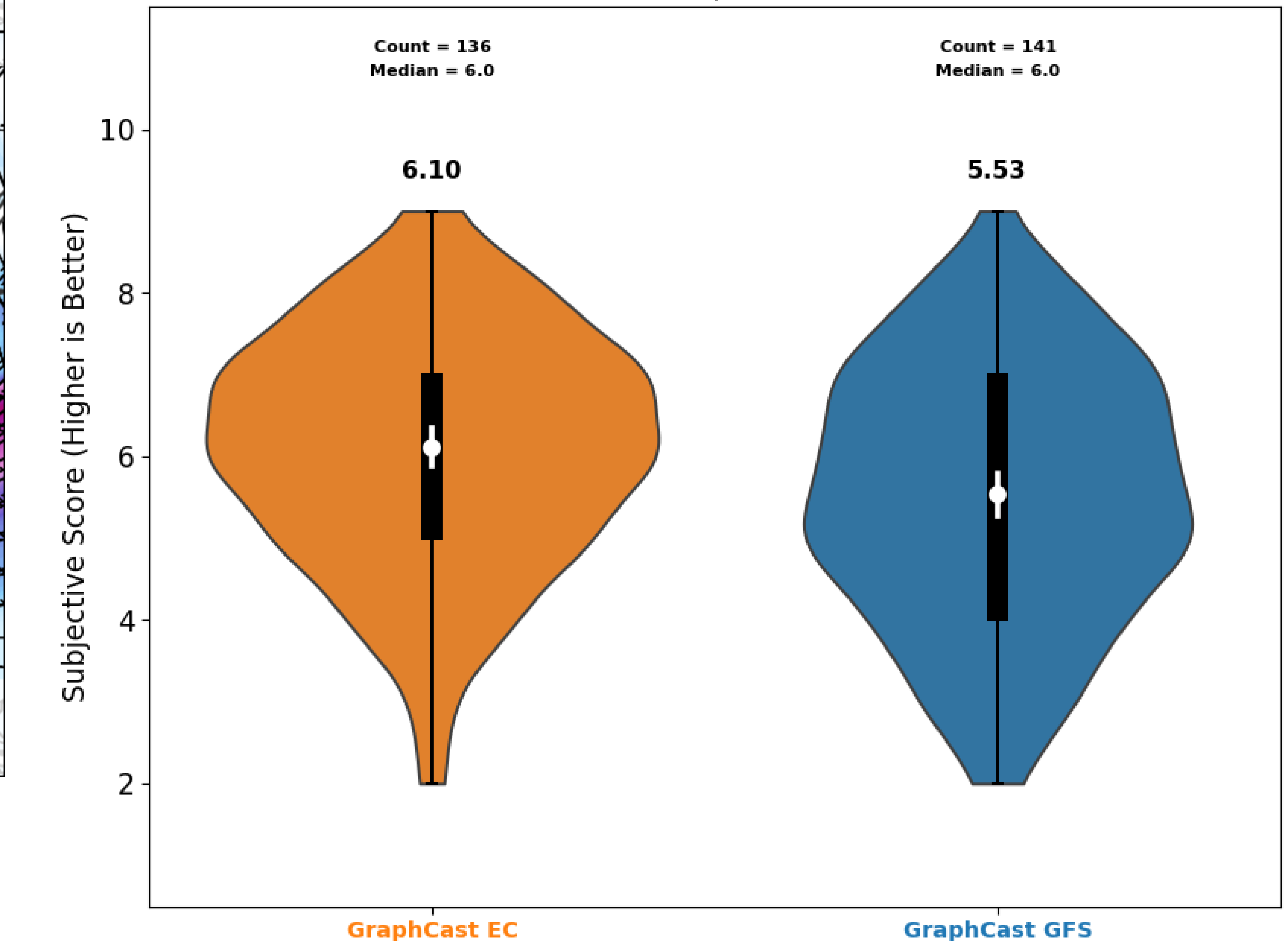


Most AI-based NWP emulators evaluated in this experiment were run in real-time by the Cooperative Institute for Research in the Atmosphere (<https://aiweather.cira.colostate.edu/>).

3. GFS vs EC ICs:

Participants directly compared Google's GraphCast with both GFS and EC initial conditions. Overall, the version with EC initial conditions was found to better capture the location and amplitude of synoptic features. The version with GFS initial conditions tended to produce weaker jets which somewhat reduced its score.

Global NWP Emulators IC Comparison: Day 7 Synoptic Pattern
All Responses



4. Emulator Ratings

Pangu-Weather was the highest-rated emulator with GFS initial conditions. All emulators except Aurora were rated higher than the operational GFS. Aurora was found to be much smoother than the other models and often produced very weak wind fields.

AIFS received the highest rating out of all the models with EC initial conditions. Most emulators were rated very similar to the operational ECMWF. Aurora continued to produce much smoother and weaker fields than the other models. Overall, models with EC initial conditions received higher ratings than those with GFS initial conditions.