Global reforecasts from MPAS "GRAF" with mesh refinement over the US and Europe

Thomas M. Hamill[1], Raghu Raj Prasanna Kumar[2], Karthik Kashinath[2], Carl Ponder[2], Mike Pritchard[2], Tao Ge[2], Akshay Subramanian[2], Jaideep Pathak[2], John Wong[1], Brett Wilt[1], and Peter Neilley[1]

[1] The Weather Company (TWC), Atlanta, Georgia USA [2] NVIDIA, Santa Clara, California, USA. For more, contact tom.hamill@weather.com

Abstract

The training of deep-learning numerical weather prediction (DLNWP) models at high spatial and temporal resolution is greatly facilitated by lengthy time series of gridded analyses or forecasts. This poster describes a reforecast data set created with the National Center for Atmospheric Research (NCAR) Model for Predictions Across Scales (MPAS) model, with some refinements made by The Weather Company (TWC) to improve accuracy and computational performance on GPUs. The MPAS model is a global model permitting variable grid spacing, in this case ~ 4 km spacing over the contiguous US (CONUS) and Europe (Fig. 1), relaxing to 15 km beyond these areas. There are five years of case days simulated, but those case days are chosen for active US weather, 2004-2024. The initial condition dates of chosen cases were selected for landfalling tropical cyclones, severe local storms, major winter storms, and heavy precipitation across major hydrologic basins. This preferential sampling should improve the ability to train DLNWP models over the CONUS for precipitation and severe weather.

The data are available in an Amazon S3 bucket. The license permits general noncommercial use immediately, and commercial use after 1 year.

Introduction

Many organizations have now successfully demonstrated global deep-learning numerical weather prediction (DLNWP) models that, in certain metrics, are competitive with or exceed the skill of reference standards from conventional numerical weather prediction systems. Increasingly, these systems are being extended from deterministic predictions to ensemble predictions, and the initially simplified state vectors are being expanded to include more relevant surface weather variables, and at higher temporal and spatial resolution.

Global reanalyses are needed at the scale of thunderstorms (say, 4 km or less, with sub-hourly archival of data), with accurate, unbiased surface data, including precipitation amount and type. Given the computational and data challenges in producing reanalyses and the strong desire for relevant training data to jumpstart convection-permitting DLNWP, we have created a reforecast data set rather than a reanalysis. This data set leverages the NCAR MPAS model, as the "GRAF" model optimized for GPUs, with mesh refinement over the CONUS and Europe. It is initialized from ERA5. It consists of 1825 cases with active weather over a 20+ year period. In the sections to follow, we provide more detail on this data set; on how initial conditions were generated, the configuration of the forecast system, the characteristics of the data that were chosen for storage.

This reforecast data set has several unique characteristics to facilitate DLNWP.

- *Five years of cases selected to emphasize storms and precipitation, spread over a 20+ year period.*
- *Forecasts +27 h, to cover spinup + full diurnal cycle. Hurricanes to + 33h*
- *Many variables saved relevant to storm forecasting, in an S3 bucket.*

Model and initial condition choices.

The TWC "GRAF" (**G**lobal high-**R**esolution **A**tmospheric **F**orecasting) is based on the NCAR MPAS model, version 6.3. In the configuration used in the reforecast, the model has these characteristics:

- 4 km grid over the CONUS and Europe, → 15 km elsewhere.
- 50 vertical levels.
- Scale-aware nTiedtke convective parameterization with near full use of the convective parameterization at 15 km to near none at 4 km.
- YSU planetary boundary layer and gravity-wave drag.
- WSM6 microphysics, RRTMG radiative transfer scheme
- NOAH 4-layer land-surface model.

Reforecast initial conditions were taken from ERA5 reanalyses at 0.25-degree grid spacing. Selection of initial-condition dates emphasized storms and precipitation. Initial dates were chosen typically 12-24 h preceding US or near-US land-falling hurricanes or significant tropical storms. With hurricanes that lingered over land such as Harvey in 2017 (Houston floods), or with hurricanes that stayed near the coast such as 2019's Dorian, or with hurricanes that had outsized impact such as Sandy in 2012 (NJ/NY), multiple reforecast dates were chosen closely spaced in time and covered more than just landfall. The only eastern Pacific hurricane date was for rapidly intensifying Otis (2023).

Many case dates were also selected based on the criteria of either observed or forecast severe weather and tornado outbreaks. We also examined the literature for dates of US predecessor rain events (PREs) associated with tropical cyclones, major northeast US snowstorms, ice storms, and west-coast atmospheric rivers, and we included a several dates for these. A synthesis of the cases selected based on these criteria are shown in Fig. 2. The rest of the cases were selected based on heavy precipitation in major hydrologic basins in the CONUS.

Finally, a few case dates were selected to fill the largest gaps between other selected case dates.

The methodological details of case date selection for hydrologic basin heavy precipitation are somewhat involved, and they are described in more detail along with more detail on the methods for case selection for other storm types in https://tinyurl.com/GRAF-refcst-config .

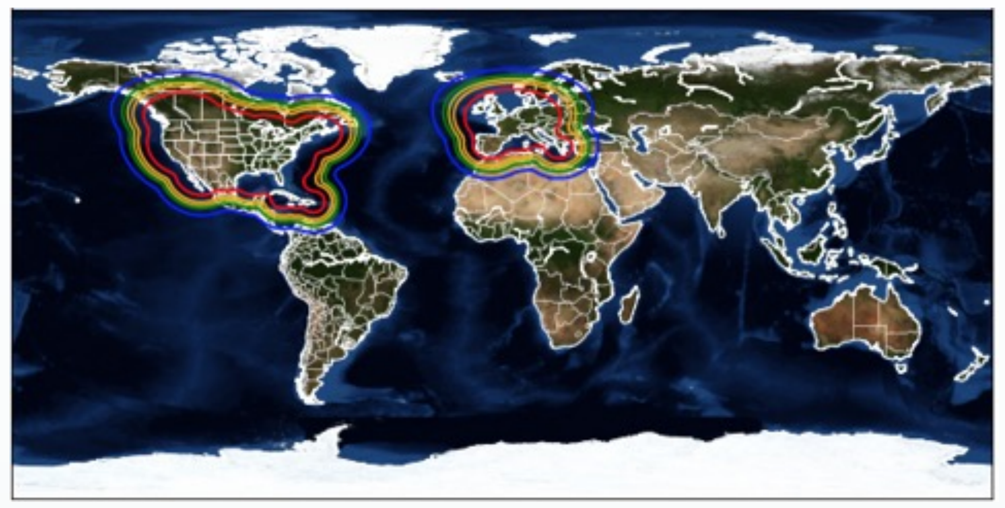

Figure 1: Variations in MPAS GRAF grid spacing. Inside the red is 4 km, outside blue is 15 km.

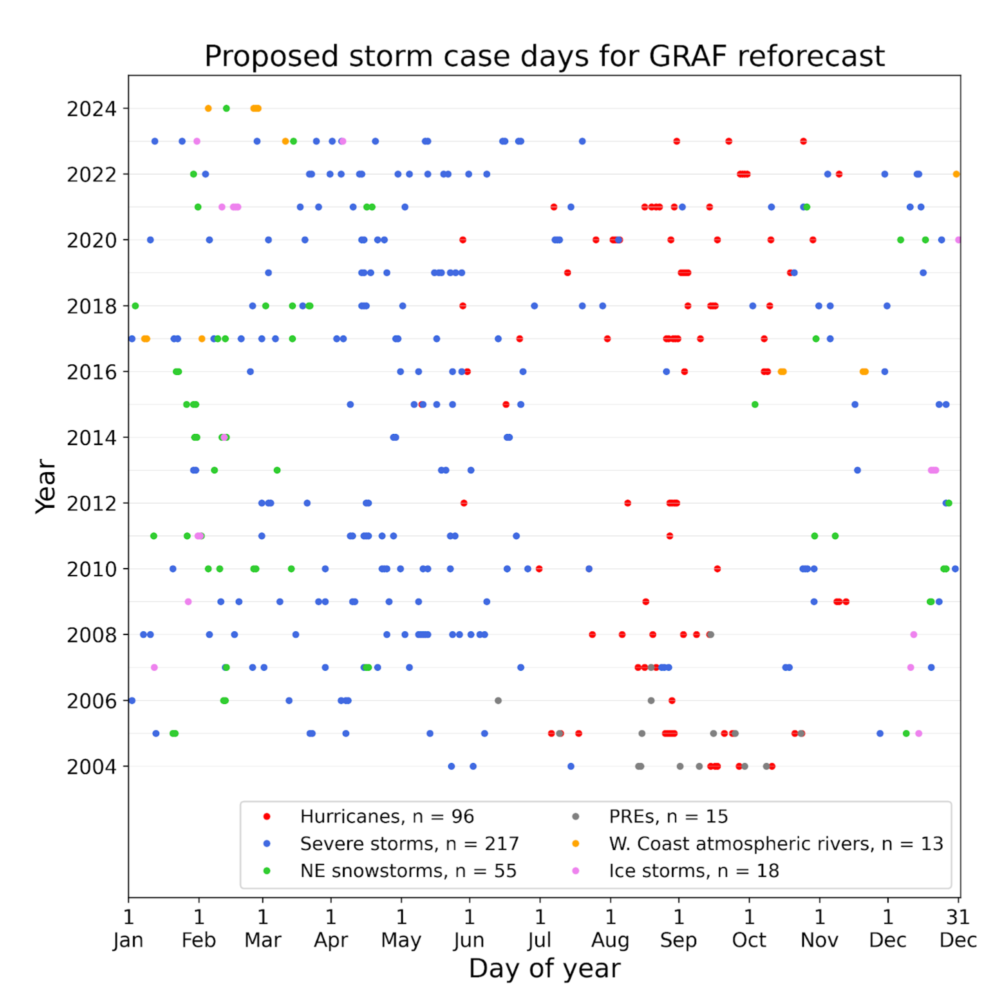

Figure 2: Initial condition dates for major CONUS high-impact weather events.

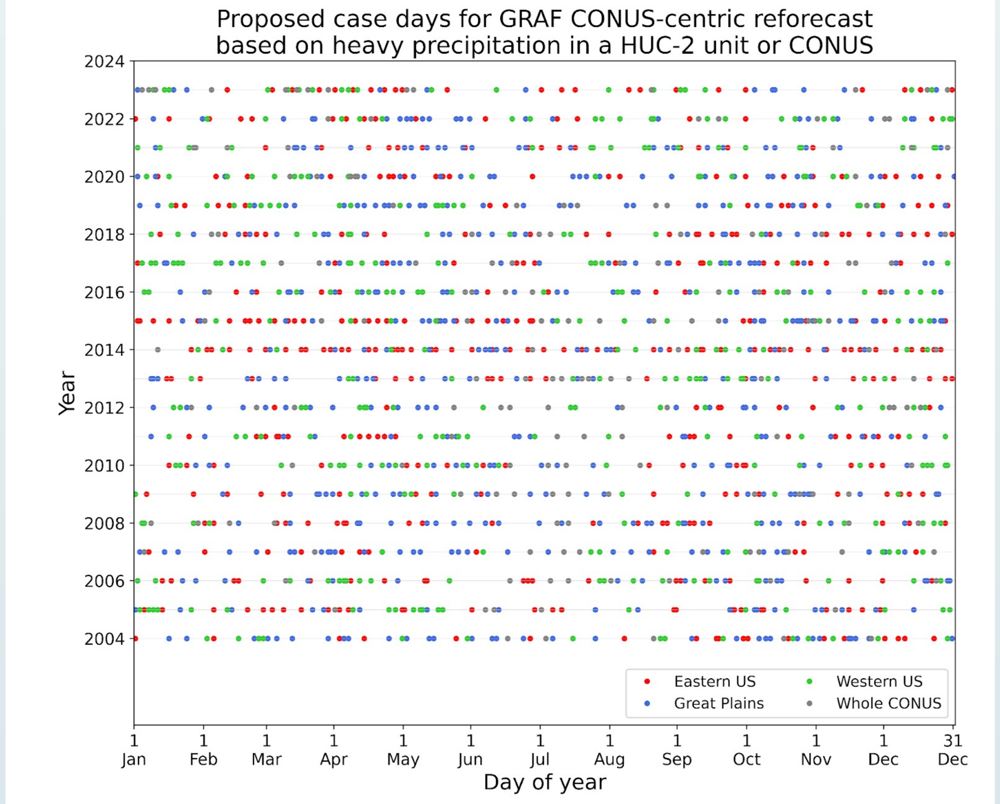

Figure 3: Initial condition dates for heavy precipitation in one of the large CONUS hydrologic units.

Output variables

Variable	Temporal frequency	2D or 3D
Cloud water (vapor, cloud, rain, ice, snow, graupel mixing ratio)	15 min	3D (atm)
u-, v-, w-wind components	15 min	3D (atm)
Total pressure	15 min	3D (atm)
Temperature	15 min	3D (atm)
Potential temperature	15 min	3D (atm)
Soil temperature, liquid water, liquid equivalent	15 min	3D (soil)
Conditional probability of rain, snow, ice	15 min	2D
10-m u- and v-wind components	15 min	2D
Downward all-sky surface flux, short and longwave	15 min	2D
Precipitation rate	5 min	2D
Precipitation type	5 min	2D
Precipitation type bucket accumulations (rain, convective rain, ice, snow, total precipitation)	5 min	2D
Total cloud cover	15 min	2D
Mean sea-level pressure	15 min	2D
2-m temperature and dewpoint, specific humidity	15 min	2D
Total-column precipitable water	15 min	2D
Skin temperature, including SST	15 min	2D
Wind gust	15 min	2D
All-sky top of atmosphere outgoing longwave	15 min	2D
Instability parameters (LI, CAPE, CIN, LCL)	15 min	2D
Ceiling above ground level	15 min	2D
Echo top (18 dBz)	15 min	2D
Fire weather index	15 min	2D
PBL height	15 min	2D
Latent heat at the surface	15 min	2D
Hourly averaged latent heat flux	15 min	2D
Power disruption index	15 min	2D
Thunderstorm potential index	15 min	2D

These are saved in chunked zarr format and will be stored in an AWS S3 bucket.

Conclusion

A reforecast data set is being created using the MPAS/GRAF model, with convection-permitting mesh refinement over the US and Europe. Data are available for non-commercial uses this year, for commercial uses next year. We hope other users find these data useful for training deep learning numerical weather prediction models.

Acknowledgements

We gratefully acknowledge support to Amazon Web Services for the free storage of these data.