HRRR Data-Assimilation System (HRRRDAS) and HRRRE Forecasts

David Dowell, NOAA/ESRL/GSL, 1 March 2020

Introduction

ESRL/GSL, NSSL, NCEP/EMC, and partners are working together on a project to design a single-model, convection-allowing, ensemble-based data-assimilation and forecasting system, called the Rapid Refresh Forecast System (RRFS). This project aims to develop advanced high-resolution data-assimilation techniques and ensemble-forecasting methods while supporting the unification and simplification of the NCEP modeling suite around the FV3 model.

While the standalone regional (SAR) FV3 model is being developed for convection-allowing forecasting, other possible components of the RRFS are being tested now in the experimental, WRF-based High-Resolution Rapid Refresh Ensemble (HRRRE). Experimental runs of the HRRRE at GSL are focused particularly on (1) improving 0-12 h high-resolution forecasts through ensemble-based, multi-scale data assimilation and (2) producing spread in 0-36 h ensemble forecasts through initial-condition perturbations, boundary-condition perturbations, and stochastic physics. The experimental HRRRE has two components -- a 36-member ensemble analysis system (HRRR Data-Assimilation System, or "HRRRDAS") and a 9-member ensemble forecast system. The HRRRDAS provides initial conditions (ensemble mean) and a background ensemble for data assimilation in HRRRv4, which is scheduled for operational implementation in June 2020. The HRRRDAS and HRRRE also provide initial and boundary conditions for a prototype Warn-on-Forecast System, which is an on-demand, regional, ensemble system that provides sub-hourly probabilistic forecast guidance for convective storms at short lead times O(1 h). This document describes the HRRRDAS and HRRRE configuration as of March 2020.

HRRR Data-Assimilation System (HRRRDAS)

The 36-member, hourly-updated HRRRDAS is the central component of the HRRR team's experimental ensemble-system development, enabling other ensemble applications such as HRRRE forecasts, HRRRv4 initialization, RTMA-3D analysis, and Warn-on-Forecast System (WoFS) initialization. Characteristics are summarized below.

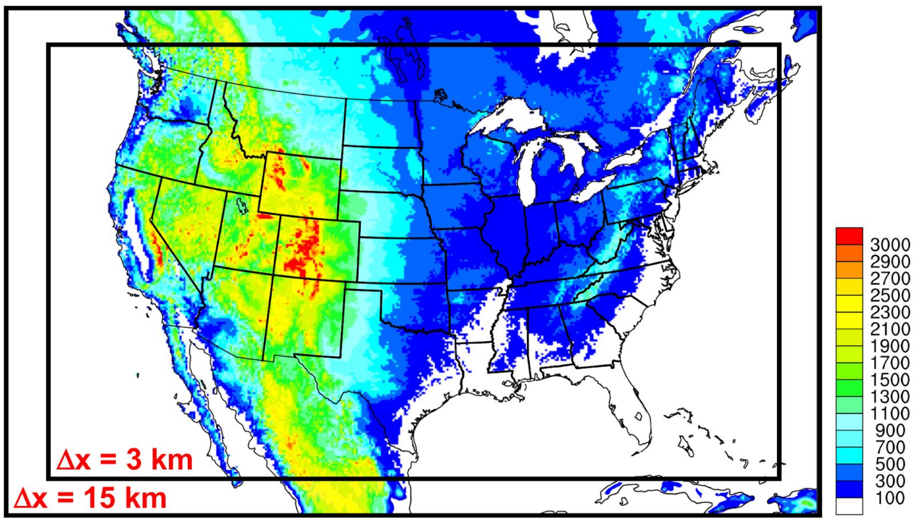

Figure 1. Terrain height (m) in the HRRRE nested 15-km and 3-km domains.

Model and Domain

- WRF-ARW version 3.9+, a combination of the public release version of WRF 3.9 plus GSL enhancements for model physics, and the same version that will be implemented operationally in HRRRv4
- Nested grids with 15-km and 3-km horizontal grid spacing
    - The scalar grid dimensions are 1799 x 1059 for HRRRv4 but are 1800 x 1060 for the 3-km HRRRDAS grid. The grid dimensions in the HRRRDAS make nesting possible (and also make the HRRRv4-HRRRDAS system a bit complicated). The 3-km HRRRDAS grid is identical to the HRRR CONUS domain except for an additional row and column on the north and east sides.
    - The 15-km outer grid exists for the purpose of lateral boundary conditions. The outer zone allows random perturbations at the boundaries of the 15-km grid to evolve before reaching the inner 3-km grid. (Previous attempts to randomly

perturb the boundaries of the 3-km domain directly produced artifacts, such as unrealistically heavy precipitation near the boundaries).

- Hybrid vertical coordinate, with 50 vertical levels, as in HRRRv3 and HRRRv4
- Physics as described in Benjamin et al. 2016, MWR http://journals.ametsoc.org/doi/abs/10.1175/MWR-D-15-0242.1 except that a deep convection parameterization is not used on the 3-km domain

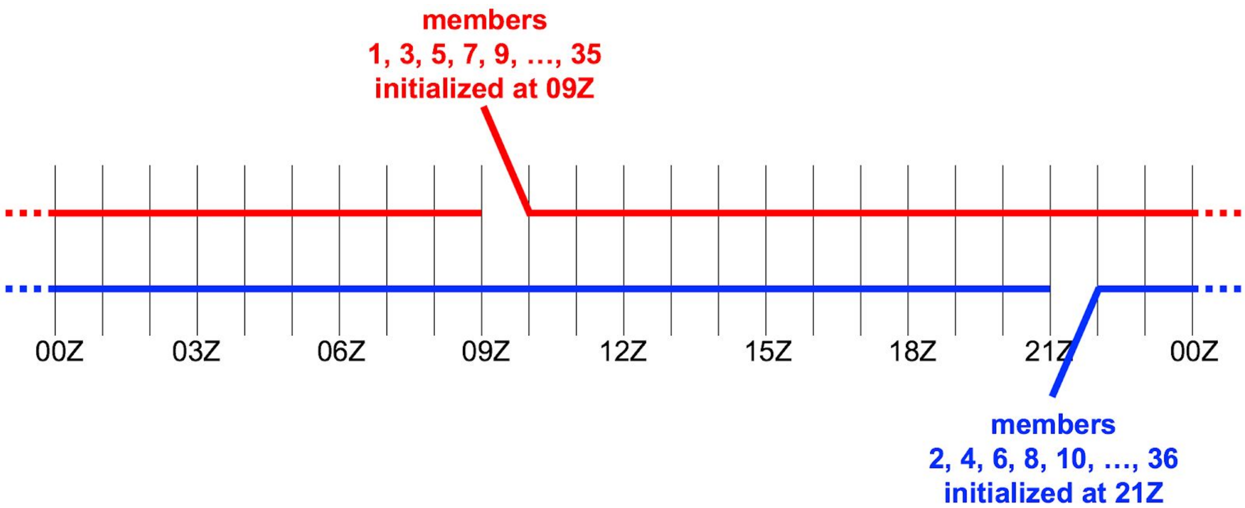

Figure 2. Daily schedule for HRRRDAS member initialization, cycling, and retirement. Thick lines represent 1-h WRF model advances. Thin vertical lines represent hourly data assimilation.

"Rolling" strategy for initializing and retiring members

- Odd-numbered members (1, 3, 5, 7, 9, …, 35) are initialized as described below at 0900 UTC and advanced 1 h to 1000 UTC, when they become part of the 36-member ensemble. These members are cycled hourly as part of the 36-member ensemble for 23 h until 0900 UTC. Afterwards, these members are retired and replaced by new members initialized at 0900 UTC.
- Even-numbered members (2, 4, 6, 8, 10, …, 36) are initialized as described below at 2100 UTC and advanced 1 h to 2200 UTC, when they become part of the 36-member ensemble. These members are cycled hourly as part of the 36-member ensemble for 23 h until 2100 UTC. Afterwards, these members are retired and replaced by new members initialized at 2100 UTC.

- From RAP analyses at 0900 and 2100 UTC, the WRF Preprocessing System (WPS) is used to create *wrfinput_d01* and *wrfinput_d02* files that represent the ensemble-mean initial atmospheric and soil conditions for the 15-km and 3-km domains.
- After the completion of WPS, the soil variables in the most recent HRRR forecast are copied to the HRRRDAS ensemble-mean *wrfinput_d02*.
- With the GSI utility program *enspreproc.x*, 36 atmospheric perturbation fields are generated from the first 36 members of the most recent GDAS 6-h or 9-h forecast. Then, the GSI utility program *initialens.x* adds these perturbations to the ensemble-mean state to create a 36-member ensemble of *wrfinput_d01* and *wrfinput_d02* files. The WRF fields that are perturbed are T, U, V, QVAPOR, and MU.
- During the first hour of WRF model integration, when a new member is introduced into the ensemble, soil conditions are perturbed through a stochastically perturbed parameterization (SPP) method.
- As indicated in Fig. 2, 18 of the 36 downscaled members are introduced into the full ensemble at any one time if the ensemble system is operating properly. However, if the ensemble system has failed, it is restarted from the full 36-member ensemble of downscaled members created at 0900 or 2100 UTC.

Boundary conditions

- Ensemble-mean *wrfbdy_d01* files are created with WPS four times per day: 0000, 0600, 1200, and 1800 UTC. At each time, boundary conditions are derived from hours 0-60 of the GFS forecast, at 3-h intervals.
- At each of the four times, 36 perturbed boundary-condition files are produced by adding random perturbations to the ensemble-mean *wrfbdy_d01*. Perturbations are drawn randomly from a "perturbation bank" that was generated with WRF-Var. The perturbed WRF fields are MU, U, V, T, and QVAPOR.
- The script that advances WRF simply grabs the most recent set of perturbed boundary conditions.

Observations

- The HRRRDAS assimilates the same hourly bufr files of conventional observations (http://www.emc.ncep.noaa.gov/mmb/data_processing/data_processing/) that RAP and HRRR assimilate. At 0000 and 1200 UTC, the "early" prepbufr file is assimilated. At other times, the standard prepbufr file is assimilated. All of these observation files are typically available 30-40 min after the hour.

- The HRRRDAS assimilates the same hourly Langley bufr satellite data that RAP and HRRR assimilate. These data are typically available 30-40 min after the hour.
- A new program *process_NSSL_mosaic_enkf* prepares MRMS radar-reflectivity observations for assimilation into the HRRRDAS. This program was based on the GSI utility program *process_NSSL_mosaic*. The new program outputs NetCDF, in the format read by GSI-EnKF versions that include the University of Oklahoma's enhancements for radar-data assimilation. The new program includes a distinction between reflectivity observations that are considered to be of precipitation and observations that indicate a lack of precipitation. The latter observations are assigned a single value (currently -10 dBZ), and the same minimum threshold should be used in GSI's treatment of the model background reflectivity. The radar preprocessor also includes observation thinning. The MRMS observations are thinned significantly (currently from ~500,000,000 raw observations to ~200,000 assimilated observations) in the HRRRDAS.

Hourly Data Assimilation

- In the first assimilation step, the ensemble mean of the 36-member, 1-h forecasts is computed and stored in the WRF NetCDF file *firstguess.mean*. Then, the GSI observer is run for this ensemble mean. The observer applies observation quality control and computes forecast innovations for both conventional and radar-reflectivity observations.
- In the second assimilation step, the GSI observer is run in parallel for each of the 36 ensemble members. In addition to computing forecast innovations, the GSI observer is used to clear clouds and precipitating hydrometeors in each member. The non-variational clearing procedure, utilizing GOES cloud top, METAR, and MRMS reflectivity observations, is the same as in the deterministic HRRRv4.
- The third assimilation step is the EnKF update of the ensemble members with conventional observations.
    - Analysis variables are U, V, T, QVAPOR, and QCLOUD.
    - The full localization radius (where weight reaches zero for the Gaspari-Cohn compact pseudo-gaussian) is 300 km horizontally and 0.5 scale height vertically.
    - After the analysis, ensemble spread is restored with relaxation to prior spread (RTPS; inflation factor 1.20 for the 15-km domain and 1.08 for the 3-km domain).

The use of an inflation factor greater than 1 is atypical but is supported by collaborative research with NCAR.

- The fourth assimilation step is the EnKF update of the ensemble members with radar-reflectivity observations, which occurs for the 3-km domain only.
  - Analysis variables are U, V, T, QVAPOR, QCLOUD, QRAIN, QSNOW, and QGRAUPEL.
  - The full localization radius (where weight reaches zero for the Gaspari-Cohn compact pseudo-gaussian) is 18 km horizontally and 0.5 scale height vertically.
  - After the analysis, ensemble spread is restored with relaxation to prior spread (3-km domain only, inflation factor 1.08).
- Our research experiments indicated that EnKF analysis of rain number concentration (QNRAIN) did not benefit the forecasts. Therefore, QNRAIN is not included in the list of analysis variables in the fourth assimilation step. Instead, in a fifth assimilation step, we specify values of QNRAIN based on the analysis values of QRAIN, according to a Marshall-Palmer type exponential distribution. This specification ensures that computed reflectivity values are realistic and that the microphysics scheme is passed realistic values for the start of the model advance. It would also be reasonable to specify cloud-water number concentration (QNCLOUD) in this step, but that is not done currently.

Experimental HRRRE Forecasts

The first 9 members of the HRRRDAS are advanced in a free ensemble forecast at selected times. The initialization times and forecast lengths can be adjusted according to testbed requests. The default forecast characteristics are as follows:

- 36-h forecast from 0000 UTC, typically available by 0900 UTC
- 24-h forecast from 1200 UTC, typically available by 1900 UTC.

In spring and summer 2020, we will run additional forecasts, to aid in WoF System initialization and to enable evaluation of time-lagged ensemble forecasts:

- Forecast from 1800 UTC, out to as many as 42 h
- Forecast from 0600 UTC, out to as many as 30 h.

The boundary conditions of the forecast members are perturbed randomly, as described in the HRRRDAS section of this document. The forecast ensemble also includes stochastically perturbed parameterizations (SPP). Currently, the Smirnova/RUC land-surface, MYNN PBL, and Thompson microphysics schemes are all perturbed continuously during the forecast.

An ensemble post-processing system is applied to the nine HRRRE forecast members to produce all-season weather hazard probabilities. HRRRE probabilities are the fraction of members that exceed a given threshold within 40 km of a point. The final probability field is smoothed using a Gaussian filter of width 25 km.

High Performance Computing

The HRRRDAS runs on kjet: Intel Skylake CPU, 40 cores/node, and 96 GB memory/node. Hourly cycling requires 108 nodes (4320 processors). Resources for the free ensemble forecasts vary according to the requirements of real-time demonstrations.

Contributors

GSL: Trevor Alcott, Curtis Alexander, Jeff Beck, Stan Benjamin, John Brown, David Dowell, Jeff Duda, Guoqing Ge, Jeff Hamilton, Ming Hu, Eric James, Brian Jamison, Isidora Jankov, Evan Kalina, Jaymes Kenyon, Terra Ladwig, Joe Olson, Tanya Smirnova, Sam Trahan, Steve Weygandt

NSSL: Pam Heinselman, Kent Knopfmeier, Thomas Jones, Pat Skinner, Nusrat Yussouf, Lou Wicker

EMC: Ben Blake, Jacob Carley, Geoff Manikin

NCAR: Glen Romine, Craig Schwartz, Kate Fossell, Ryan Sobash, May Wong

OU: Yongming Wang, Xuguang Wang

Acknowledgments

Development of convection-allowing ensemble-based data-assimilation and forecasting capabilities is supported by the Joint Technology Transfer Initiative (JTTI) through the grant "Advancing frequently-updating storm-scale ensemble data assimilation and prediction towards operations". We appreciate the significant contributions from the high performance computing team at GSL to this and related projects. Additional computing time on jet during demonstration periods was provided by the Hurricane Forecast Improvement Project.