

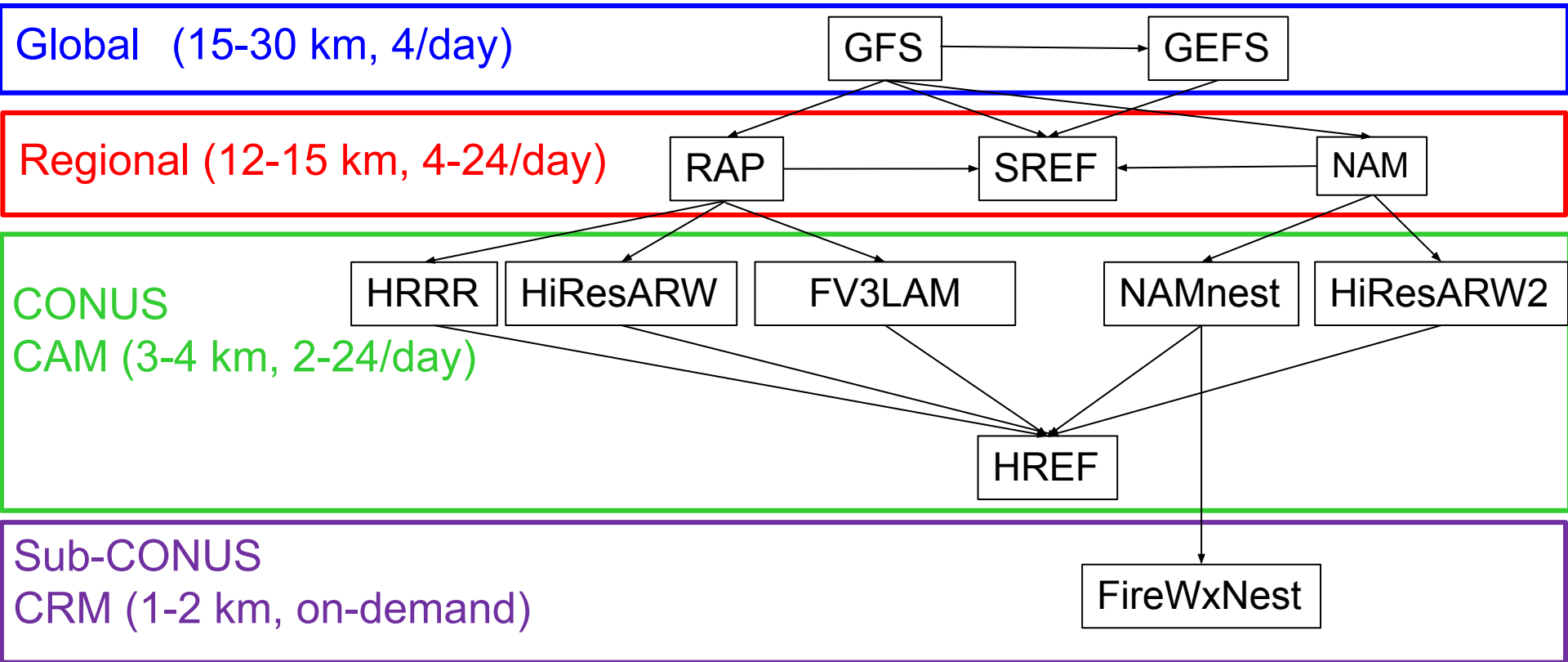
# High Resolution Ensemble Development

Curtis Alexander (NOAA/OAR/GSL)  
Jacob Carley (NOAA/NWS/EMC)

24 August 2023

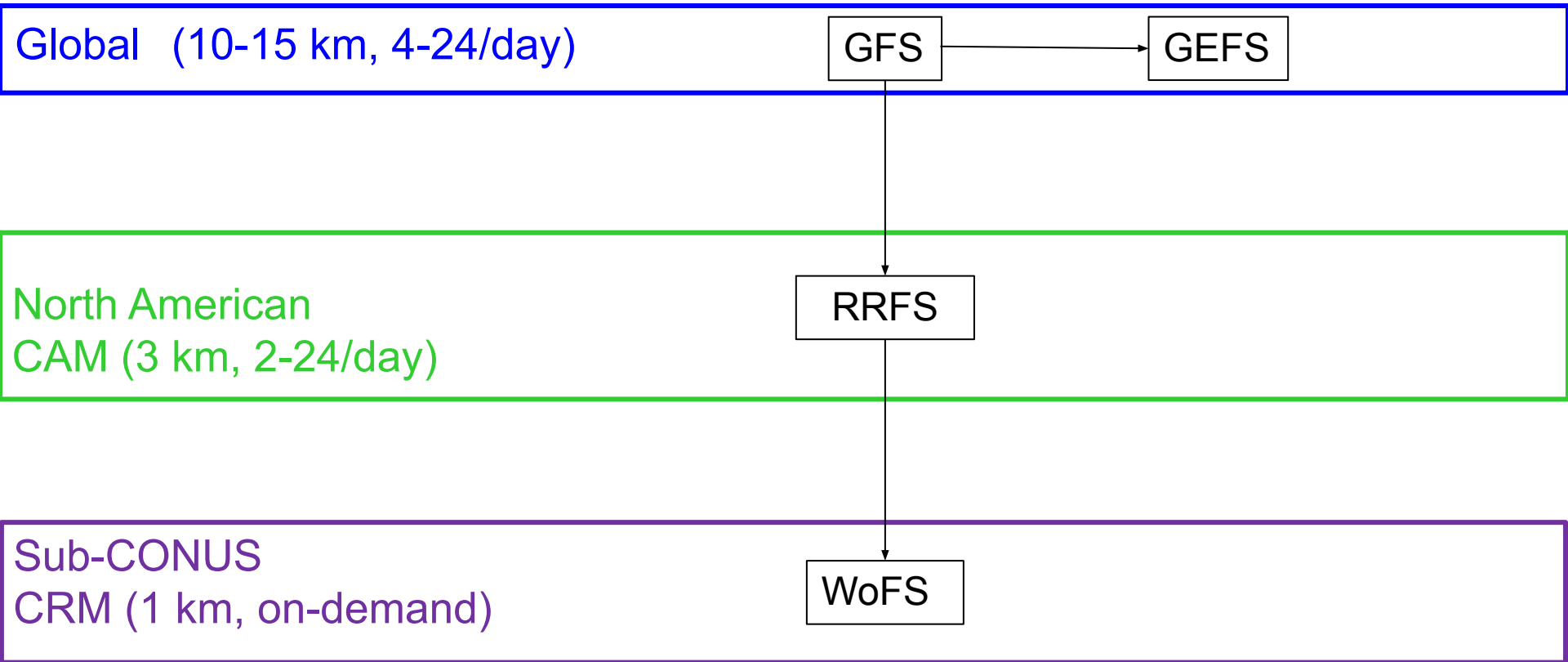
# Current Production Layout

FY23



# Future Production Layout

FY2?



## High Resolution (CAM) Ensemble Design Characteristics

How important are theoretical design principles in ensemble construction? Priorities of forecast length/updates?

Existing (baseline) HREF ensemble is highly irregular in design but highly effective ensemble prediction system

- Comprised entirely of control members (multi-model, multi-physics, multi-IC/BC, time-lagged)
- Often provides sufficient spread but is uncontrolled sampling the forecast PDF
- Spread derived, in part, from disparate, but known, control member biases
- Drives considerable complexity to the regional operational production suite that is not sustainable
- Retirement of existing HREF ensemble components becoming increasingly urgent given legacy codes
- HREF has reached its performance ceiling and additional development not tenable

RRFS ensemble design will be considerably different from HREF but needs competitive spread-skill

- Comprised of at least some perturbed-IC/BC/physics members
- Initial configuration may still contain multi-model (HREF members), multi-physics, multi-IC and/or time-lags
- Gradual progression towards a more formal ensemble design of a control plus perturbed members
- Ensemble would eventually sample forecast PDF in a more systematically controlled manner
- Establish equal likelihood of all (perturbed) members that would speak to the (un)certainly of control solution
- Ensemble product generation needs to carefully consider inclusion/exclusion of control member

Regardless of underlying member construction there is stakeholder tension of forecast lengths/update frequency

- Current HREF ensemble coverage to 48 hrs with requests to 84 hrs (day 3)...when should GEFS step in?
- Significant resource trade-space between update freq (down to hourly), forecast lengths and ensemble size
- Initial RRFS with O(5) concurrent members, update freq and lengths varied by member

## High Resolution (CAM) Ensemble Visualization & AI/ML/Calibration

CAM ensembles are expensive in compute, storage and dissemination – how can we improve efficiency in use?

- Need to retain ability to interrogate individual (all?) members
- Recent study from WoFS team showed that among forecasters examining WoFS output, the individual member product viewer was used most often (WoFS has 18 forecast members)
- Value-added products are nice (bias corrected) but raw outputs are preferred

How useful is ensemble mean (probability matched or otherwise) for high-resolution ensembles? Do we need a central state that looks like a member?

- Ensemble mean for continuous fields will almost always beat individual members but not really meaningful for non-continuous fields like hydrometeor and other cloud related outputs (PMM is useful)
- Traditional mean is helpful for quick look at the environment and pattern - but not if forecast is bimodal
- Contour band depth can help to identify the central state

Given the overall expense of CAM ensembles - how can we produce enough (fixed?) ensemble forecast data for effective AI/ML/post-processing applications? Decades of retros from CAMs are untenable due to expense.

- Some work on applying transfer learning to carry forward past training results to new model/versions - reduces some demand for very large initial data sets
- A CAM ensemble w/ same dycore & physics may allow for fewer retro runs as all members have the same climatology (GEFS does this). Multi- model or -physics will require more runs and thus is more expensive.
- Run reforecast once per day to cover retro period - this may not fit on the production machine (cloud?)

## High Resolution (CAM) Ensemble Verification and Product Generation

CAM ensembles have an extremely diverse set of stakeholders – how best to evaluate their performance?

- DTC metrics workshop (2021) established a priority list of forecast fields and measures - leveraged in construction of EMC Verification System
- Statistical comparisons of ensemble performance (e.g. reliability, ROC, etc...) need significance testing
- METplus etc...need to deploy bootstrapped or related significance information with sample size estimation
- Documented ensemble biases need to be shared/communicated with stakeholders
- Product consistency of uncertainty estimation between various predictive scales will need focus - by whom?