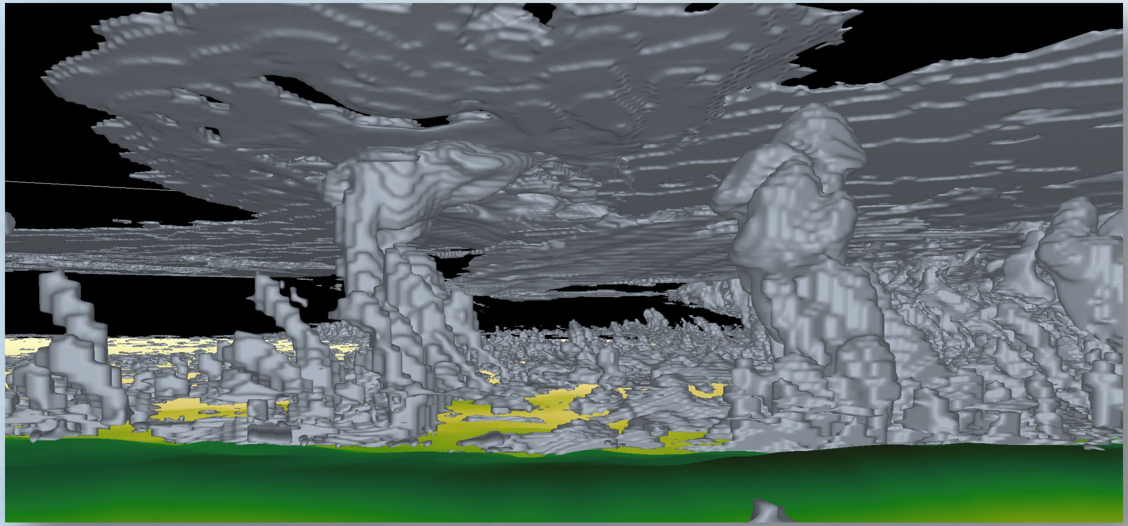
Community Session: LASSO-CACTI Scenario for Deep-Convection with Large-Eddy Simulation



28 July 2021

William I. Gustafson Jr. [1], Andrew M. Vogelmann[2], Satoshi Endo[2], Tami Fairless[2], Adam C. Varble[1], & Heng Xiao[1]

[1] Pacific Northwest National Laboratory, [2] Brookhaven National Laboratory

https://www.arm.gov/capabilities/modeling/lasso

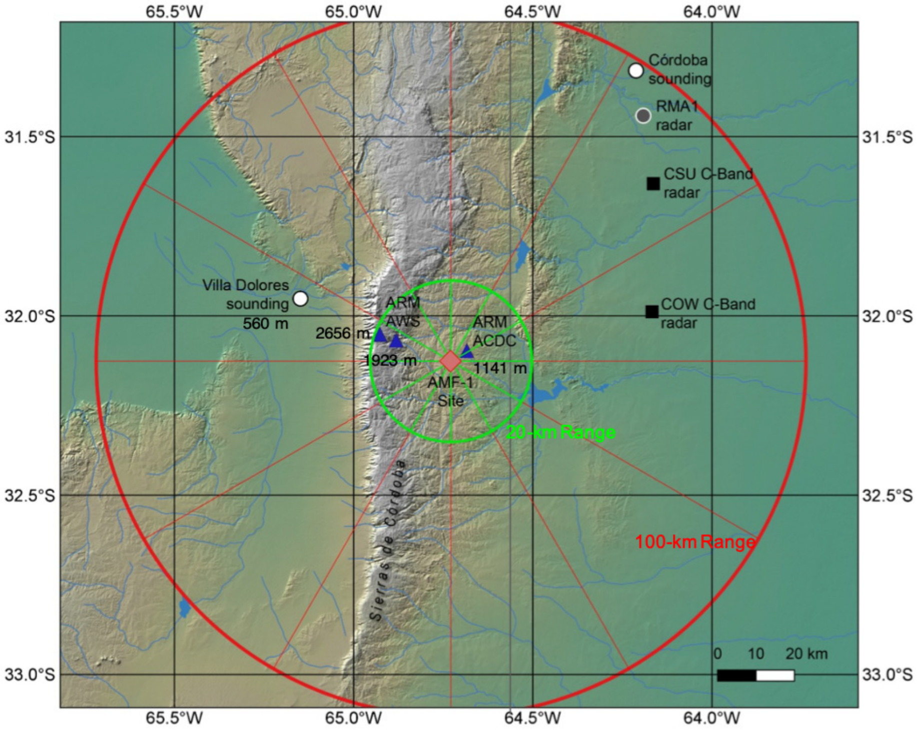
Agenda for LASSO-CACTI session

1. Introduction
2. Modeling strategy for mesoscale and LES runs
3. Model output strategy for frequency and variables
4. Strategy for observations, diagnostics, and skill-scores
5. Case selection
6. User access and tools for working with LASSO-CACTI data

What is LASSO-CACTI?

- LASSO = LES ARM Symbiotic Simulation and Observation
- LASSO seeks to add value to ARM observations by using high-resolution modeling to bridge scale gaps and add context to observations
- The CACTI field campaign occurred in 2018–2019 in Argentina with a focus on large-scale convection and its upscale growth
- LASSO will use large-eddy simulation (LES) to simulate ~10 CACTI cases with results released in 2022

Map of CACTI Deployment in Argentina



Science drivers guiding scenario design

- Convective cloud dynamics
  - e.g., thermal-like structures, updraft strength, and entrainment; the relationship to critical features like updraft and downdraft mass fluxes, vertical transport, and the shallow-to-deep convective transition
  - Convection-environment interactions, e.g., cold pools
  - Convective drafts in turbulent flow
- Microphysics-dynamics interactions
  - Especially in the context of cloud-scale eddies and smaller-scale turbulence

- Science drivers chosen to balance relevant science with computational capacity
  - LES resolution governed by cloud core requirements
  - Domain size determines portion of lifespan simulated
  - Limiting ensembles to mesoscale simulations with the potential for a small number of LES ensemble members for specific cases
  - Focusing on ~10 cases with varying convective behavior

What we seek from you today

- Input on any coarse corrections prior to starting full production LES runs
  - Does the modeling strategy meet the anticipated research needs?
  - What is the right balance of cases vs. ensemble members?
  - Are we providing the right variables from the model? Output frequencies?
  - Are the proposed dates good?
- Is the “less bundled” approach to the observations appropriate?
- Are there other ways you would like access to the data?
  - Are there obstacles to the tools you would use to analyze the runs?

Modeling Strategy

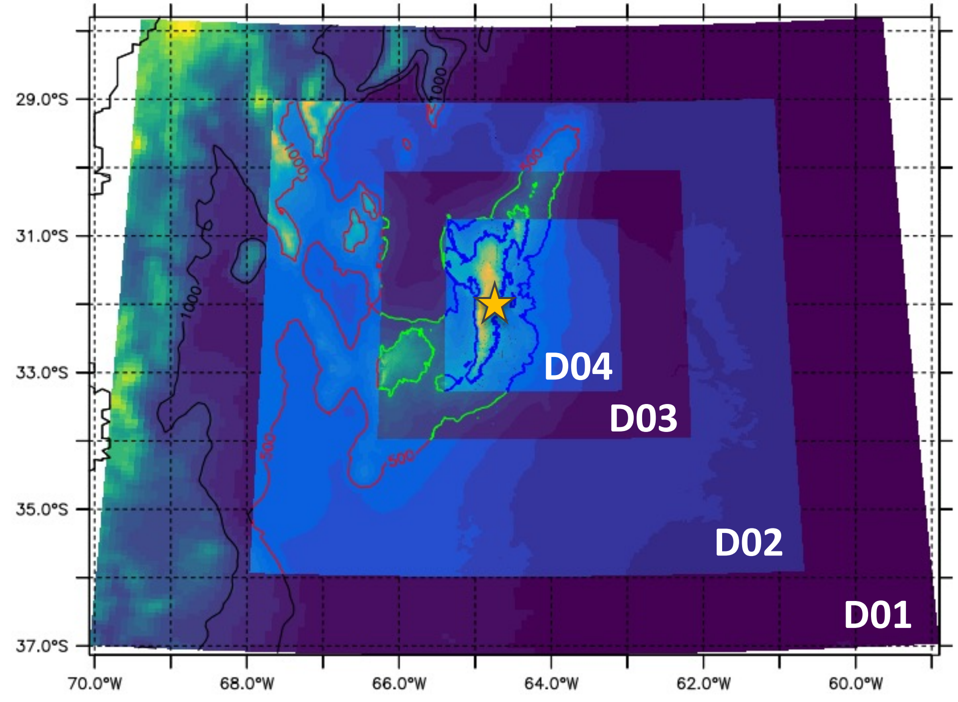
Modeling stages: goal of Δx=100 m for deep convection

- **Stage 1:** Mesoscale ensembles with Δx=7.5 & 2.5 km
  - For selecting boundary conditions and case selection
- **Stage 2:** LES setup with Δx=500 & 100 m
  - For selected cases, possibly several LES per case
- **Stage 3:** Post-process data to simplify usage

Domain Sizes

Δx =	D01 7.5 km	D02 2.5 km	D03 500 m	D04 100 m
N_x	130	258	750	2145
N_y	136	306	865	2775
	Meso		LES	

Potential WRF Domains
Δx = 7.5 km, 2.5 km, 500 m, & 100 m



★ = AMF location

Input data

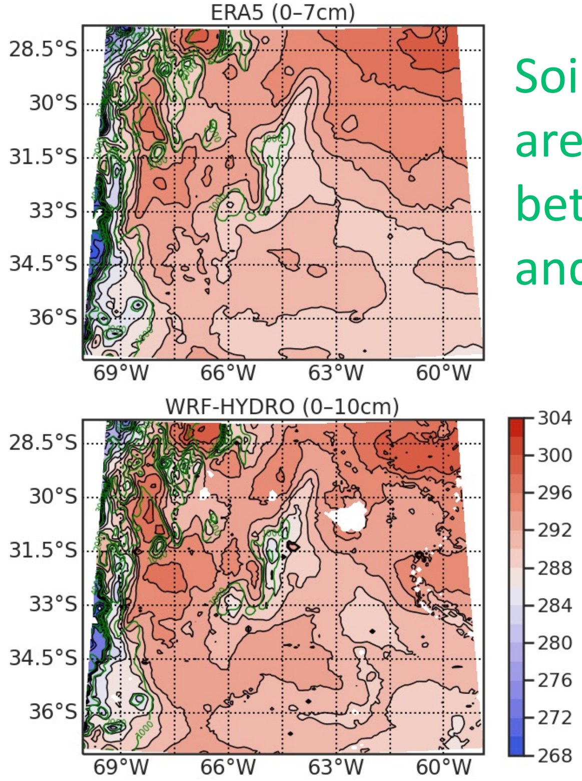
- Using MERIT DEM data for terrain elevation (Yamazaki, GRL, 2017)
  - Raw data at 3”, Δx ~ 90 km at equator
  - Smoothing for model stability using ~1 km spatial scale
- Soil initialization
  - Current prototyping runs use soil data from the host boundary condition dataset
  - Testing use of WRF-Hydro to establish a spun-up soil state consistent with WRF physics

Soil comparison for 26-Oct-2018 0 UTC:

ERA5

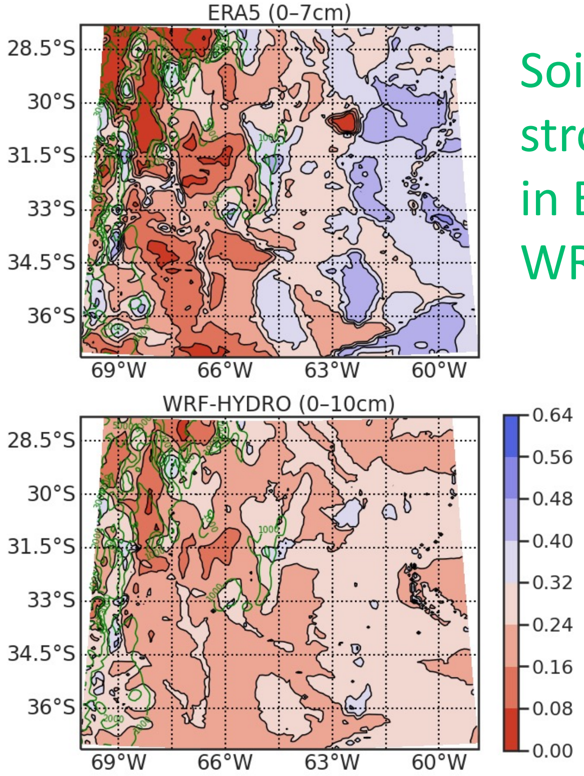
WRF-Hydro

Topmost Soil Layer Temperature [K]



Soil temperatures are similar between ERA5 and WRF-Hydro

Topmost Soil Layer Moisture [m^3/m^3]



Soil moisture has stronger gradient in ERA5 versus WRF-Hydro

Model physics configuration

- Basic physics setup is a derivative of WRF's "CONUS" physics configuration

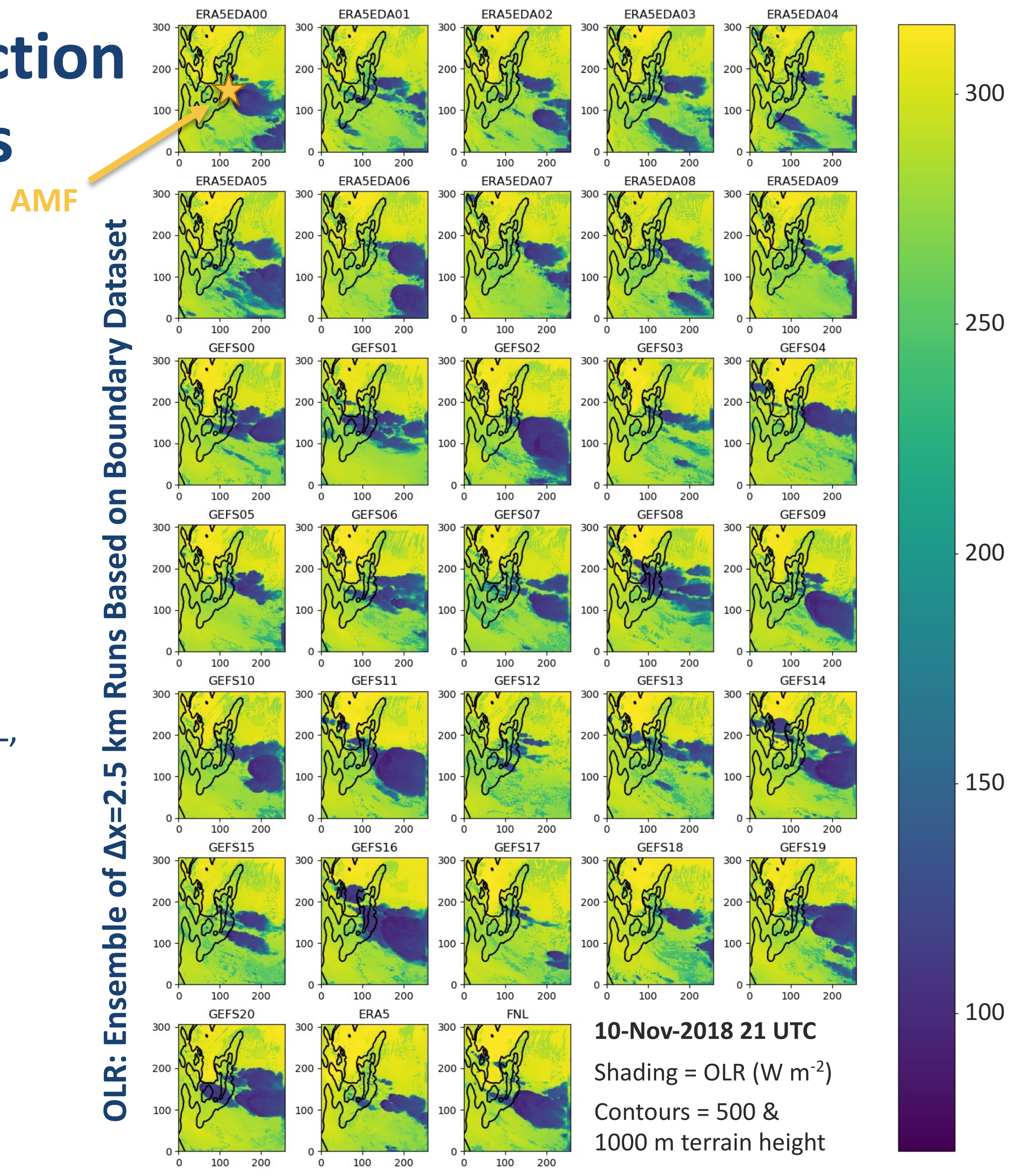
Physics Option	Number	Name
mp_physics	8	Thompson Microphysics
cu_physics	6	Modified Tiedtke Cumulus (only Δx=7.5 km)
ra_lw_physics	4	RRTMG Longwave Radiation
ra_sw_physics	4	RRTMG Shortwave Radiation
bl_pbl_physics	2	Mellor-Yamada-Janjic TKE PBL (only Δx=7.5& 2.5 km)
km_opt	2	1.5 Order TKE SGS (only Δx=500 & 100 m)
sf_surface_physics	2	Noah Land Model (considering Noah-MP)

- Considering use of aerosol-aware Thompson with climatological aerosol field

Mesoscale ensembles for case selection and LES boundary condition choices

- Selecting LASSO-CACTI cases
  - We aim to release LES for about 10 case dates
  - Selection of dates driven by convective initiation near the AMF site
  - Down-selection involves using mesoscale ensembles to test boundary condition data

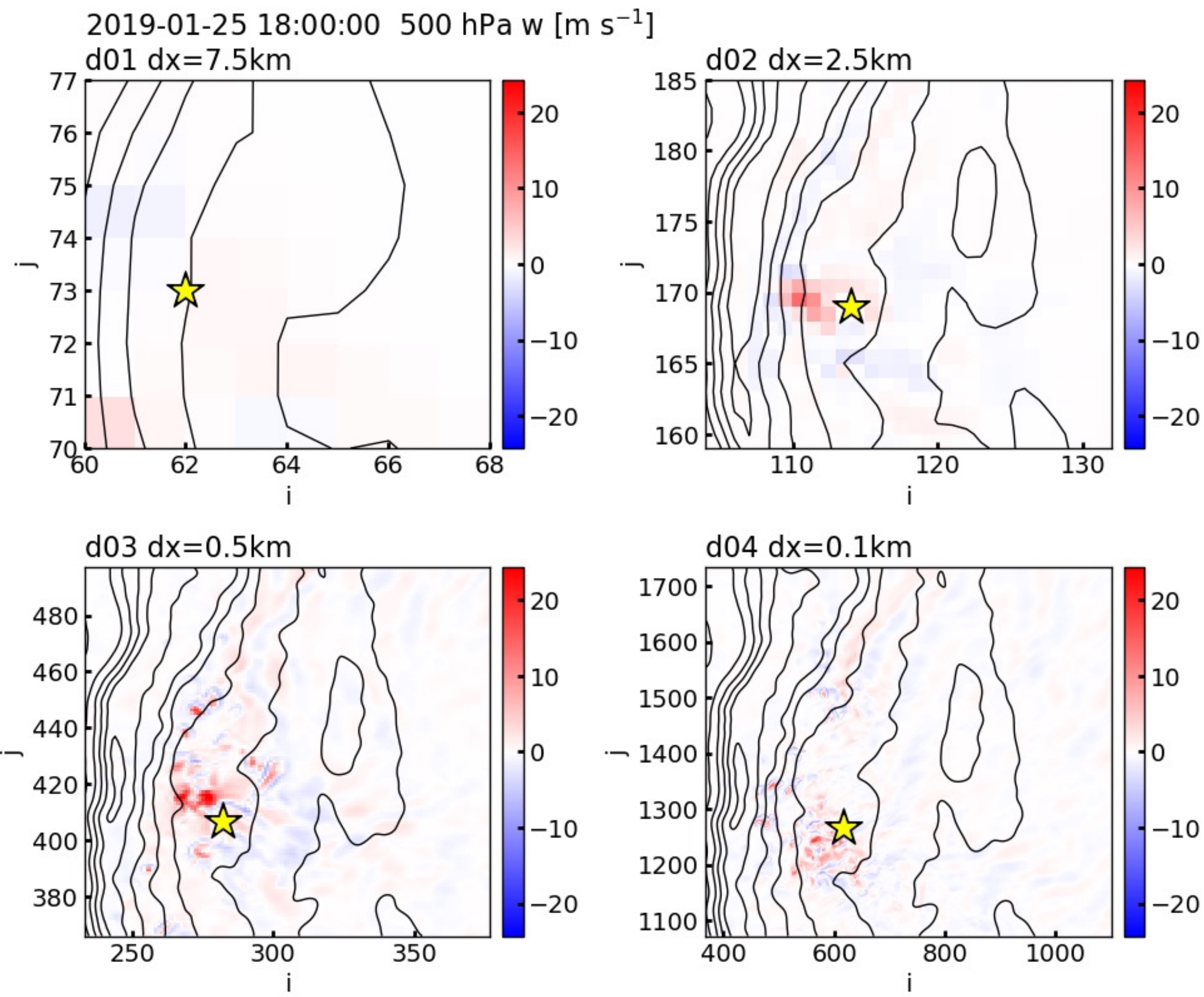
- Ran mesoscale ensembles for 20 candidate case dates (example for 10-Nov-2018 at right)
  - 33 ensemble members based on ERA5, ERA5 Ensemble, FNL, and GFS Ensemble
  - Nested down to 2.5 km grid spacing
  - Best performing ensemble members identified based on cloud comparison to GOES-16 IR data
  - Down-selected ensemble members get final vetting using bulk CSPAR2 statistics, e.g., height of Zh echo-top height at 40 dBz



LES domains

- "Ndown" from D02 to D03
- Nesting permits starting domains at different times to save resources
  - D03 generally starts at 0 UTC
  - Depending on case, D04 can start at 6 or 12 UTC
- Primary LES run based on best-performing mesoscale ensemble member(s)
- ~17 h wall time per model hour on 3200 cores, about 2.5 weeks for total run (default output, 4x per h)
- Possibility of additional LES runs for testing other BCs or microphysics

Comparison of 500 hPa Vertical Velocity at Each Grid Spacing



Feedback on the modeling strategy

- Does the modeling strategy meet the anticipated research needs?
- Thoughts on using aerosol-aware Thompson with default climatological aerosol field?
- What is the right balance for number of ensemble members and sensitivity tests to model physics/setup?
- Are there any issues we should further consider?

Model Output Strategy

Summary of model output strategy

- Output will vary depending on domain
- Will provide both raw WRF output and post-processed subsets
- ***See the list of output variables available in the XLS file shared with this session***
- Mesoscale ensembles will have 33 members, LES will have 2–3 per case date
- We expect the total dataset for the scenario to exceed 1 PB
  - Total size will depend on how we post-process the data for adding grouped subsets of variables

Raw output from WRF

- "Category" refers to a specific output file and its output frequency
  - For example, wrfout_d01 maps to Meso1
  - LES2 would be an extra "auxiliary" file with different variables from domain D04

Category	Domain(s)	Δx	Frequency	Period	Purpose
Static	All		N/A	Time 0	Static variables
Meso1	D01, D02	7.5 km, 2.5 km	15 min.	Entire run	Full model state and diagnostics
Bridge1	D03	500 m	15 min.	Entire run	Full model state and diagnostics
LES1	D04	100 m	5 min.	Entire run	Full model state and diagnostics
LES2	D04	100 m	1 min.	POI	Detailed process studies for important portion(s) of run
LES3	DO4	100 m	<=10 s	POI	Offline entrainment calculation; short period of time run later from restart file(s) based on results of full run
Restart	All		30 or 60 min.		Doing restarts; determine frequency based on time to compute as well as output these files

POI = Period of Interest

Subsets generated in post-processing

- Goal of subsets is to reduce size of files for users who do not need the whole raw file
- Variables within each subset group based on a related theme:
  - Static data, e.g., terrain height and other variables that only need to be output once
  - Meteorological state (with staggered variables moved to cell centers)
  - Meteorological state for staggered variables
  - Cloud data
  - Surface data
  - Boundary layer data
  - Radiation data
  - Tendency data (including microphysics tendencies/process rates)
  - Radar simulator (if we run CR-SIM)

Questions

- How useful to you would the microphysics tendencies/process rates be? Would you use them?
  - 22 variables based on code from MC3E CRM Intercomparison; specifics in the associates XLS file
  - Output at the coarsest time frequency, i.e., 15 min. (maybe 5 min. for D04 LES)

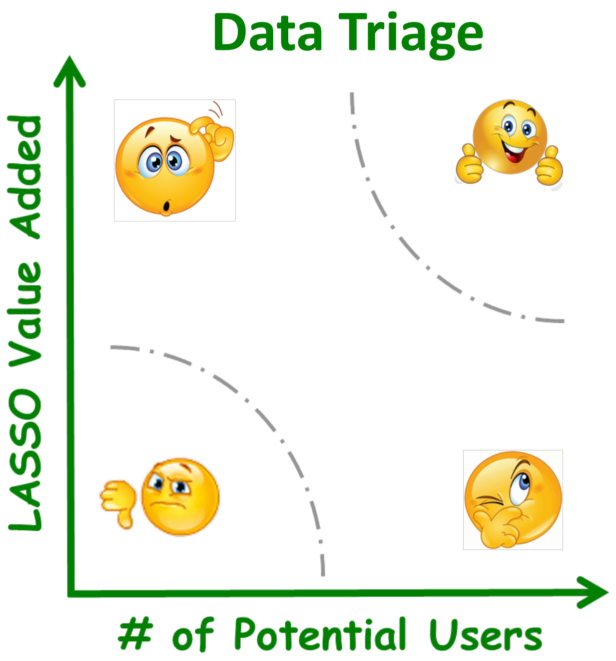
accumulated warm rain	graupel effective radius	ice formation number from liquid freezing	conversion liquid to ice/snow by riming
accumulated ice rain	condensation/evaporation of droplets	deposition/sublimation rate	raindrop radar reflectivity, lamda=10 cm
raindrop effective radius	evaporation of rain	drop freezing rate	snow reflectivity, lamda=10 cm
cloud droplet effective radius	conversion of droplets to raindrops	drop melting rate	graupel reflectivity, lamda=10 cm
cloud ice effective radius	change in number due to conversion of droplets to raindrops	conversion ice/snow to graupel/hail by riming	
snow effective radius	primary ice nucleation rate	conversion liquid to graupel/hail by riming	

- How important is it for all files to be fully readable by WRF-Python library?
- Are we missing anything for the output strategy?

Strategy for Observations, Model Diagnostics, and Skill scores

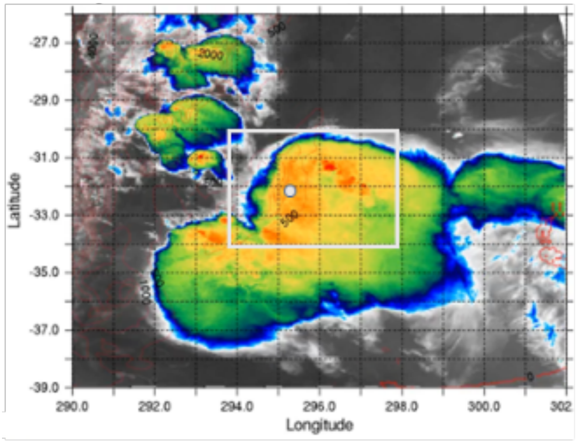
Objectives

- Approach to quantitatively evaluate model output with CACTI observations
  - Assess model setup/configuration from sensitivity tests
  - In operations, identify promising mesoscale ensemble members for further use
  - Communicate *mesoscale and LES quality* through simulation skill scores and diagnostic plots
- Purpose is to facilitate use as much as is possible, but LASSO is not conducting “PI science”
  - Largely automated, not manually intensive
  - Tried and true methods – “bulletproof”, not experimental
  - Data triage based on broad use and LASSO value added



Multiscale Observational Datasets

- **Regional: Satellite-based**
  - Sources
    - VISST: IR brightness temperatures (11.2 µm channel)
  - Application
    - Time-dependent areal coverage of the convective cores (< 219 K)

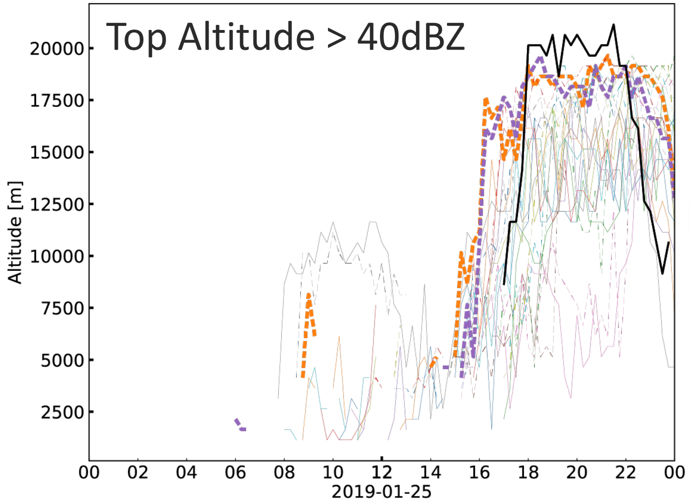


- **Local: Scanning C-Band Radar-based**
  - Sources
    - CSAPR-2/Taranis, RELAMPAGO C-Bands (TBD)
  - Applications
    - Locate AMF-storm position within the LES grid
    - Time series of surface rain rates, and of radar echo-top heights (varied dBz)
    - Diagnostic plots of Radar Contoured Frequency by Altitude Diagrams (CFADs)
    - To be discussed: PPI & RHI diagnostic plots using a radar instrument simulator, CR-SIM (Oue et al., 2020)
- **Point Measurements**
  - Sondes (thermo, ARM & RELAMPAGO)

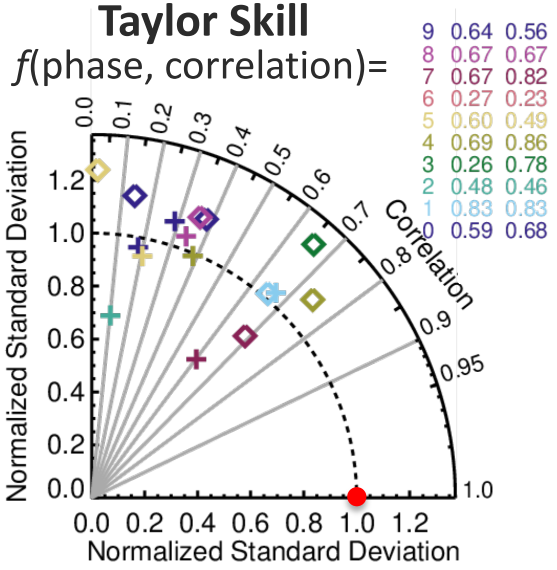
Simulation Diagnostics and Skill Scores

- Diagnostic plots to assess CACTI simulations with observations
  - Time series
  - Phase space relationships for relative relationships between a set of variables (e.g., CFADs)
- Simulation skill scores
  - Based on the Taylor diagram skill and relative mean
  - RMSD, when Taylor skill not possible (i.e., when no activity yields only 0's)

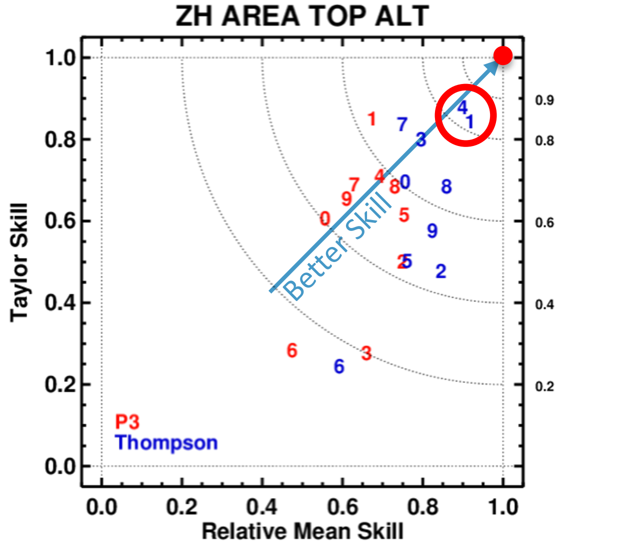
Time Series Plot



Taylor Skill
f(phase, correlation)=

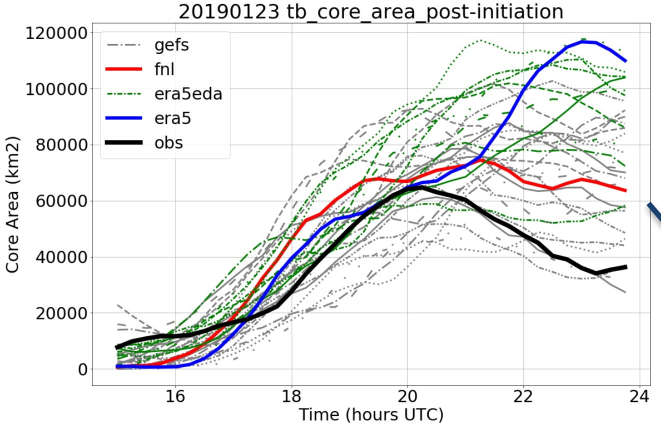


Taylor Skill & Relative Mean

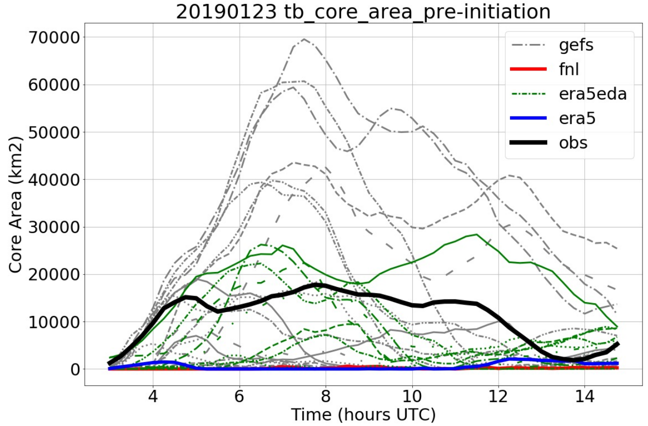


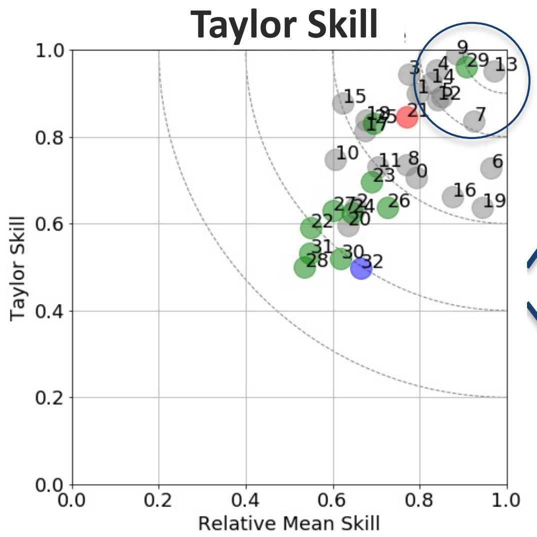
Convective Core Area Skill Score Example: 1/23/2019

Post-Initiation (15-24 Z)

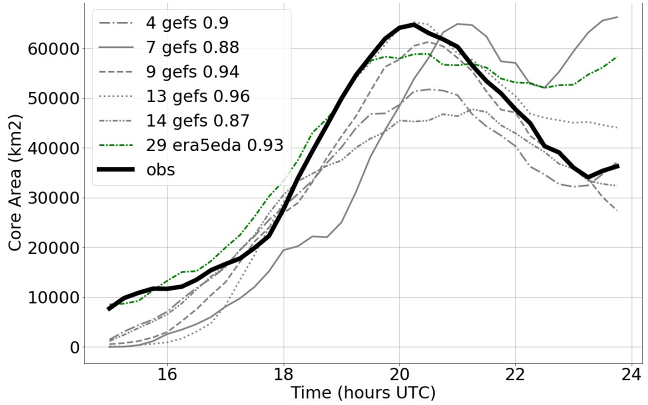


Pre-Initiation (3-15 Z)

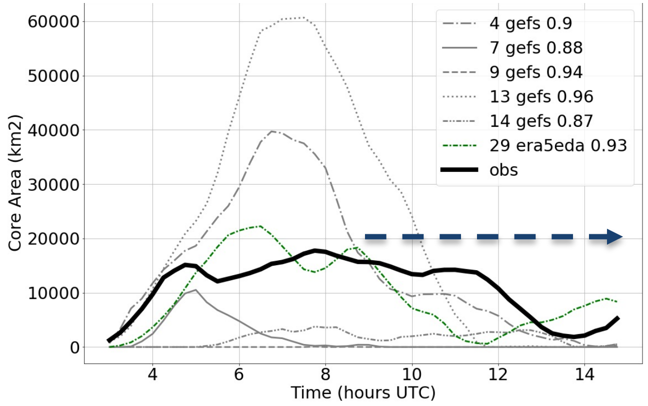




Post-initiation curves based on top 6 post-initiation skill scores



Pre-initiation curves based on top 6 post-initiation skill scores



Evaluation Plan: Skill scores & Diagnostic plots

- **Mesoscale skill scores – Ensemble vetting**
  1. Time series performance of the convective core area growth IR brightness temperature metrics (post-initiation, pre-initiation)
  2. Manually verify hourly, 2D brightness temperature snapshots agreement “near” the AMF
  3. Time series performance compared to CSAPR2 Zh echo-top height at 40 dBz (also other dBz’s?)

- **LES skill scores**
  4. #1-3 from above
  5. Regional rain rate time series (may require an AMF “location correction” within the LES grid)

- **LES diagnostic plots**
  5. Soundings from network
  6. CR-SIM instrument simulator comparisons? (to be discussed next)

Discussion Items

- **How valuable are instrument simulator comparisons of radar reflectivity (i.e., CR-SIM)?**
  - Comparisons are storm-AMF location dependent so potentially labor intensive/~PI-level effort
  - If useful, what?
    - CFADs?
    - Standardized plots of RHIs or PPIs?
- **Will the planned skill scores and diagnostics plots meet your needs?**
  - Anything missed? (e.g., Are horizontal winds at cloud base very useful?)
- **LASSO-CACTI data bundling approach**
  - We will provide processed data used in the skill scores
  - Similar to LASSO-ShCu, we do not plan to provide data that we did not process (e.g., C-Band)
  - Would this "less bundled" approach limit how you would use the data?
    - If yes, what would you like to see?

Case Selection

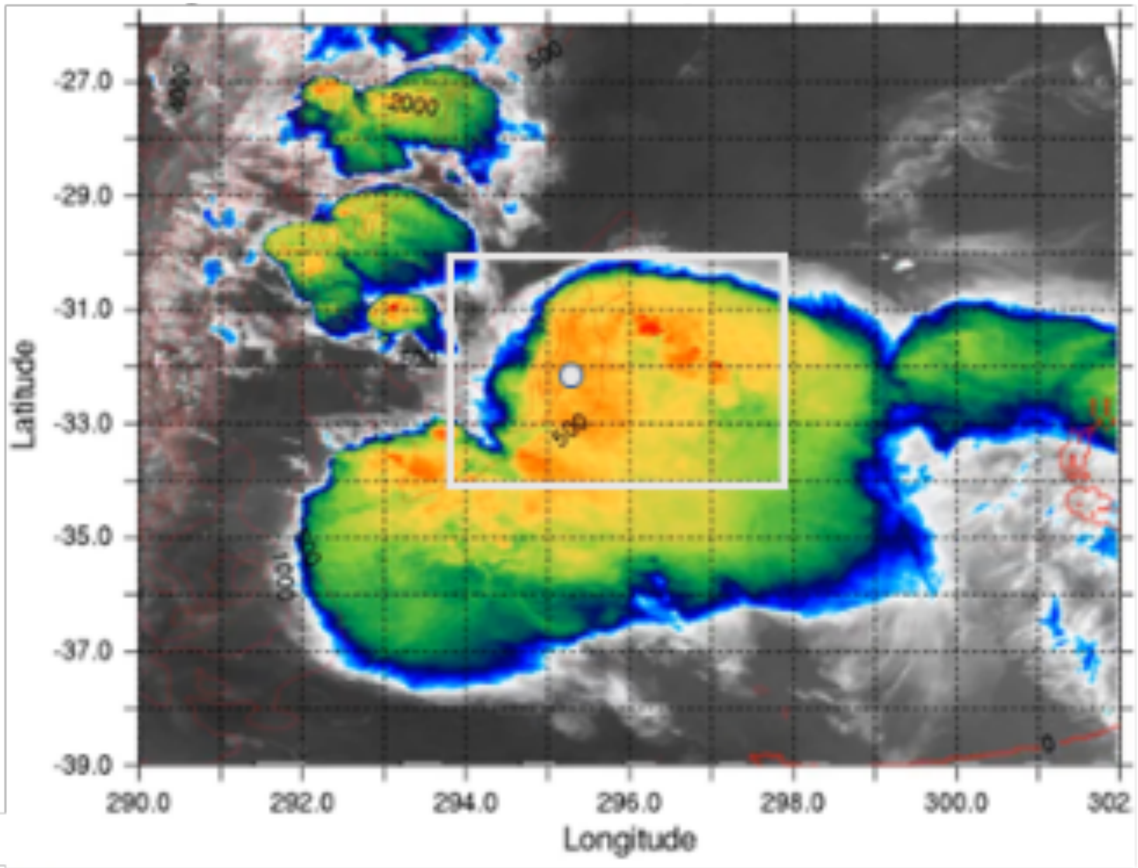
Case selection process

- **Goal:** Choose about 10 case dates that have convective initiation and growth near the AMF
  - Cases should have some degree of variability in the style of convection
  - Cases prioritized based on likelihood of success and available observations

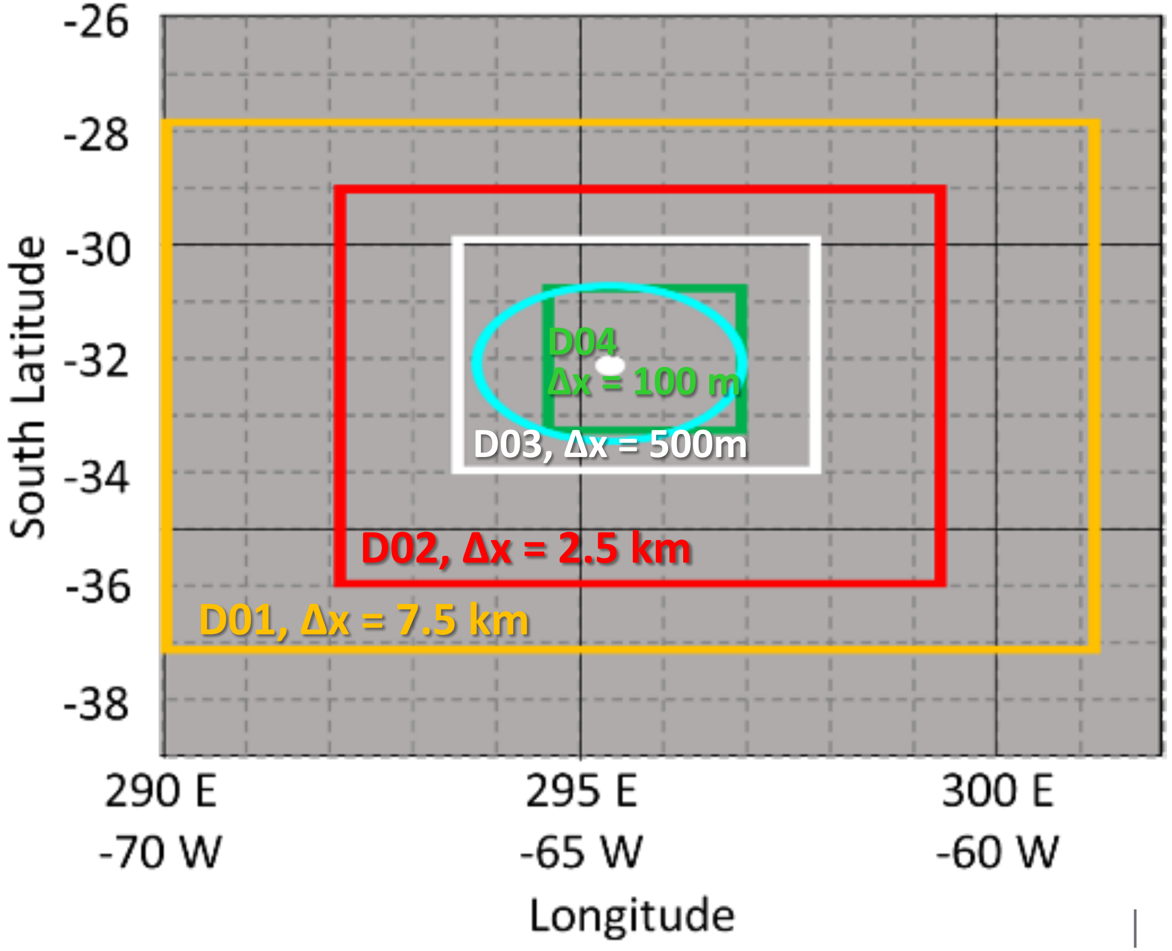
- **Approach:**
  1. Started with 20 days identified by Adam Varble, CACTI PI, that had convective initiation near the AMF
  2. Used VISST IR satellite data to evaluate convective behavior within the CSAPR2 domain
     - Eliminated days where:
       - Convection was only outside the CSAPR2 domain
       - Initiation happened outside the domain and then clouds propagated into the domain
  3. Used mesoscale ensemble with Δx=2.5 km to evaluate potential boundary conditions for each date
     - Days where all ensemble members struggled have a lower priority

- **Result:**
  - 10 potential case dates: 4 primary, 3 secondary, 3 stretch cases

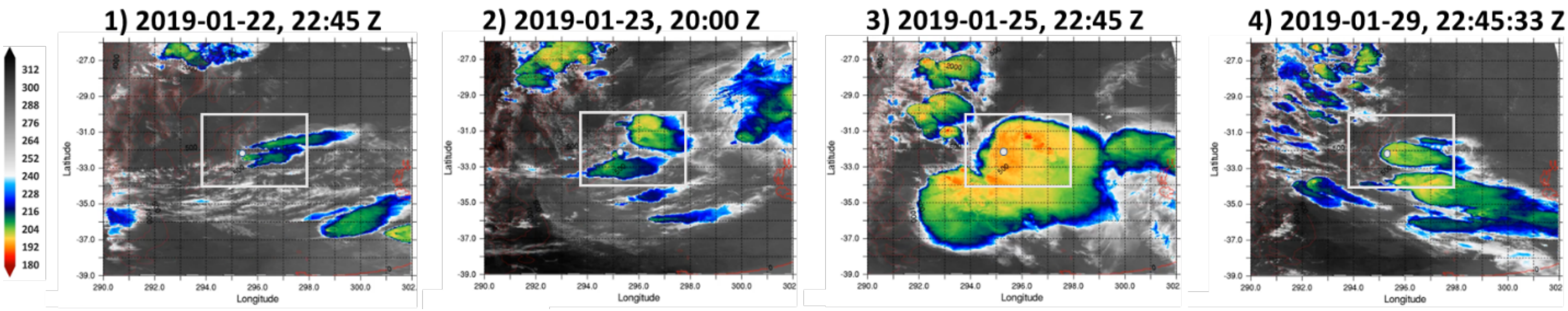
Reference grids for viewing satellite images



Rough Comparison of
CSAPR2 (ellipse) & Potential WRF Domains (rectangles)



Primary cases: Plan to simulate these days



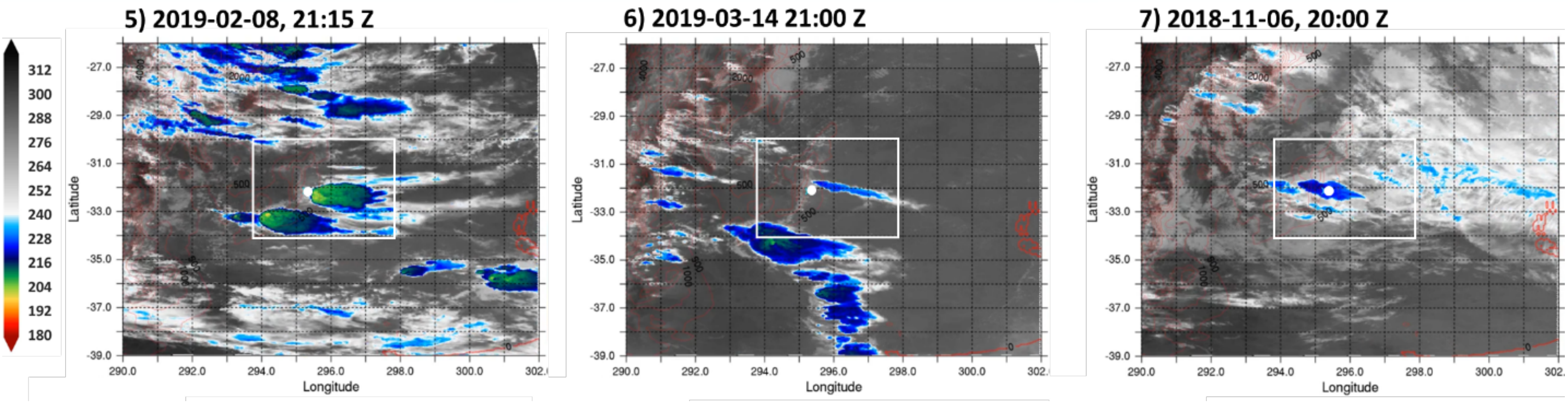
1. 2019-01-22: Small-scale, isolated convection with a second cell developing at the end of the day. Strong shear is present.

2. 2019-01-23: More vigorous convection occurring the day after the first case. Shear is less pronounced than before. Potentially it may be more challenging to simulate because the very deep development/pop occurs from the interaction and subsequent merging of two small systems within the study region.

3. 2019-01-25: The most vigorous convection of all cases. Two large systems merge within the study domain to generate a monster system.

4. 2019-01-29: Example of a mostly isolated deep convective system that develops on its own. It is more vigorous than on 1/22. Shear has increased again by this day.

Secondary cases: Listed in tentative priority – Want input

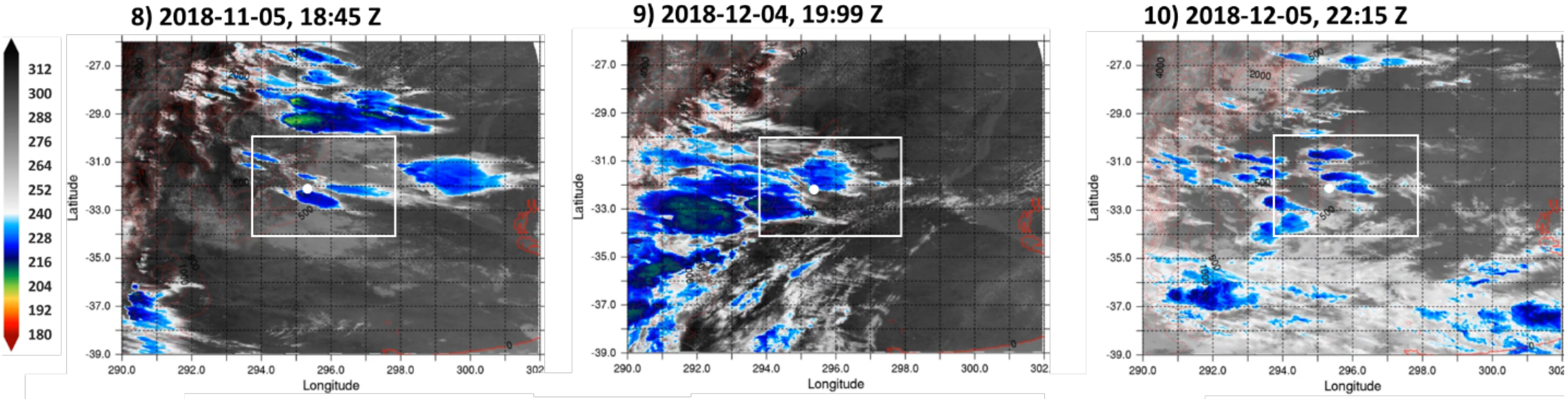


5. 2019-02-08: Multiple initiations occur within the Δx=500 m study region, which then merge into a deeper system. All mesoscale ensembles capture the phase but underestimate the magnitude of the upscale growth, possibly because important fine-scale features need Δx=100 m for an improved overall simulation. CSAPR2 went down during the event

6. 2019-03-14: Weak, isolated convection in a strongly sheared environment. Does not blow up as in the prior cases. CSAPR2 only did W-E HSRHIs.

7. 2018-11-06: Weak system develops that does not go deep, similar to 2019-03-14, but drops off within the day *so part of the system decay is captured within the domain*. This case may be harder to simulate because there are multiple, small convective cells that form and dissipate within the local region prior to the study event that might not be captured in the large-scale forcings, although some ensemble members still seem promising.

Stretch cases: May or may not be simulated – Want user input for prioritization (or removal)



8. 2018-11-05: 2–3 small initiations occur within the domain and grow to merge, but they get overrun at 21Z by systems advecting in, so the simulation is of most use before then. Two or three of the mesoscale ensemble members seem viable.

9. 2018-12-04: System develops within the domain north of the AMF. The case has a lot of complicated activity, and all mesoscale simulations dramatically underestimate the growth of the core area, but that might be improved at Δx=100 m?

10. 2018-12-05: Scattered, small-scale convective initiation but none of them grow. Only two ensemble members look potentially viable. Is there scientific interest in this case?

Discussion: 10 Candidate cases

- **4 Primary cases**
  1. 2019-01-22: Small-scale, isolated convection
  2. 2019-01-23: More vigorous convection
  3. 2019-01-25: Monster convection
  4. 2019-01-29: Isolated deep convection

Input desired as to ***whether the number of simulated ensembles should be increased*** for a particular case of special interest.

- **3 Secondary cases**
  5. 2019-02-08: Multiple initiations merge into a deeper system.
     CSAPR2 went down during the event.
  6. 2019-03-14: Weak, isolated convection.
     CSAPR2 only did W-E HSRHIs.
  7. 2018-11-06: Weak system develops that does not go deep.
     More complicated atmosphere than on 3/14.

Input sought for reprioritizing based on potential scientific benefit.

- **3 Stretch cases**
  8. 2018-11-05: 3 small initiations grow/merge; get overrun by 21Z.
  9. 2018-12-04: All ensembles underestimate growth of a medium system.
  10. 2018-12-05: Scattered, small-scale convective initiation but none grow.

Input sought for reprioritizing based on potential scientific benefit *or removal*.

Any other observational issues we should be aware of for these dates?

User Tools and Data Access

We want to make it as easy as possible for users, but…

- Using these runs will be non-trivial due to the data size!
  - Mesoscale ensemble for D02 (w/o subsets)
    - ~160 GB per ensemble member
    - >100 TB for full set of cases and members
  - LES runs for D04 (w/o subsets)
    - Raw output >50 TB per run
    - >1 PB raw model output for 10 cases & 2 LES/case
    - Restarts, subsets, and sub-minute output would be in addition to this

Rough File Sizes for Each Domain

Δx =	D01 7.5 km	D02 2.5 km	D03 500 m	D04 100 m
N_x	130	258	750	2145
N_y	136	306	865	2775
Snapshot Size	0.3 GB	1.1 GB	18 GB	154 GB

NOTE: Sizes are *rough* estimates and likely low depending on final decisions, e.g., do not include MP tendencies

Access to the data

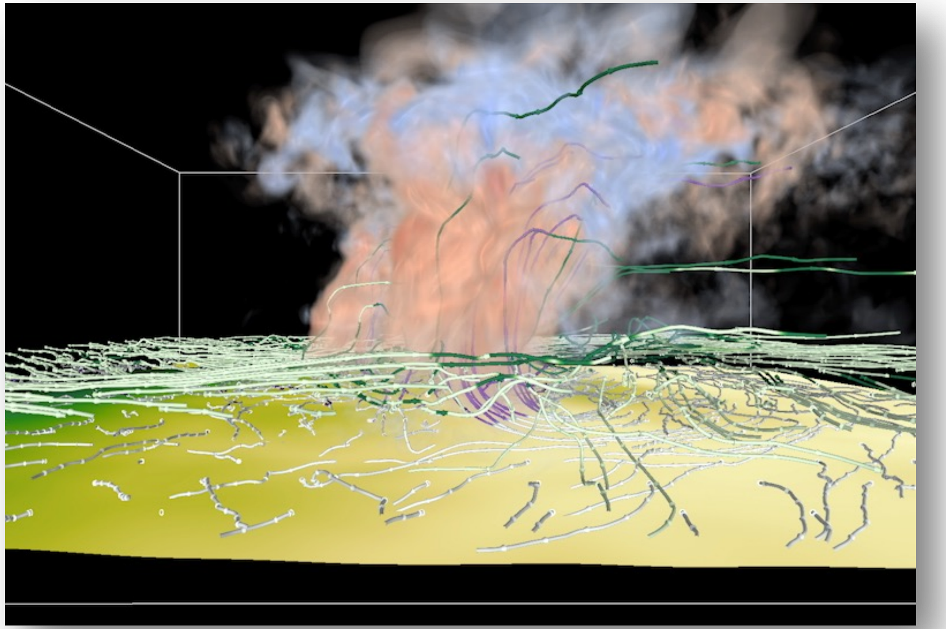
- Option 1: Download the data
  - A reasonable choice if one works on other DOE HPC and use Globus for the transfers
  - We will provide a web interface to simplify the ordering process

- Option 2: Work on ARM's cluster
  - Planning to maintain a copy of at least the more commonly used data on Cumulus at ORNL
  - Can request an account from ARM
  - Two methods
    - Python Jupyter Notebook: Will provide a few examples to help users get started with basic analysis and plotting of the LASSO-CACTI data
    - Direct login with access to job submission queue: For the more discriminating user who requires more than the Notebook interface can offer

Discussion:
What software tools would you want available?

- Jupyter notebooks with typical Python modules
- HPC environment for direct login
  - Intel compilers and associated modules for MPI, netCDF, etc.
  - Anaconda tools for Python environment
  - Job management via SLURM
  - Resources available to users will be subject to ARM's infrastructure needs, e.g., data processing, LASSO production
- Paraview for 3-D visualization
  - Able to use client-server approach to make remote visualization possible

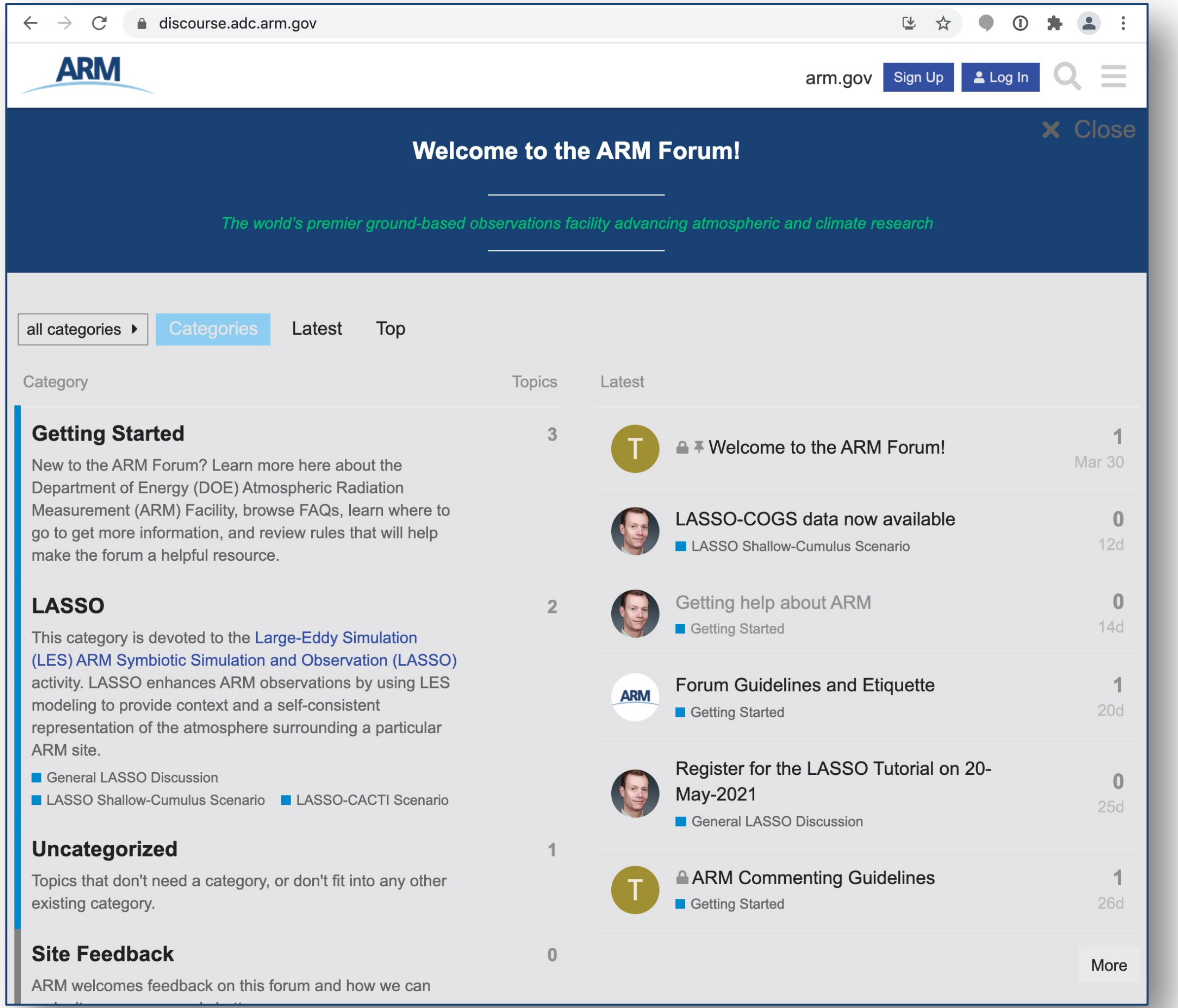
WRF, Δx = 100 m
Vertical Velocity of Cloud Core Region and Streamlines, 25-Jan-2021 20 UTC



Shading: Red=W Up; Blue=W Down
Streamlines: Seeds at 2 km AMSL (white-to-purple) and 5 km AMSL (light to dark green)
Produced with VAPOR software from NCAR

Join the community! New online forum for LASSO

- Check out the new online forum for LASSO: https://discourse.adc.arm.gov/
- Use it for user support, discussing scenario development, and related topics around LASSO and ARM
- Aiming for it to become an online resource for LASSO information and support
- Other ARM topics besides LASSO are also possible—ask us if you would like a category added, e.g., for a field campaign or value-added product

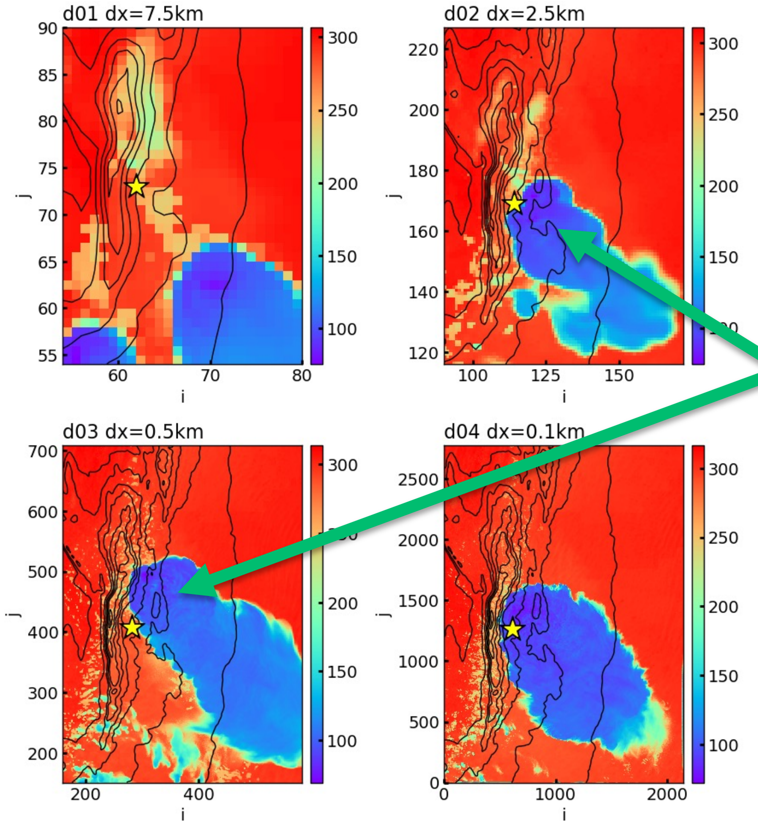


Extra Slides

Mesoscale vs. large-eddy simulations

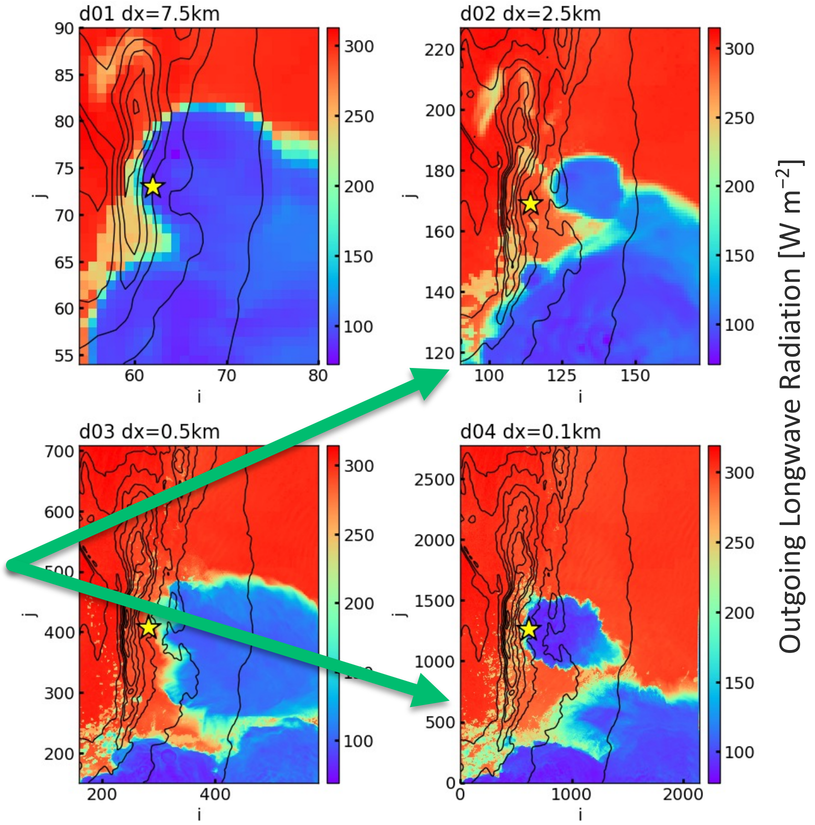
Finding that mesoscale simulations are only semi-predictive of cloud development within LES

WRF's OLR, 25-Jan-2019 20 UTC
Forcing = GEFS Member #1



- Substantial variability between ensemble members with mesoscale grid spacings (d01 & d02) necessitates careful choice of boundary conditions
- Location of convective development shifts along ridge between grid spacings
- Size of cloud system varies between grid spacings

WRF's OLR, 25-Jan-2019 20 UTC
Forcing = GEFS Member #2



Contours = Terrain height, 300 m interval