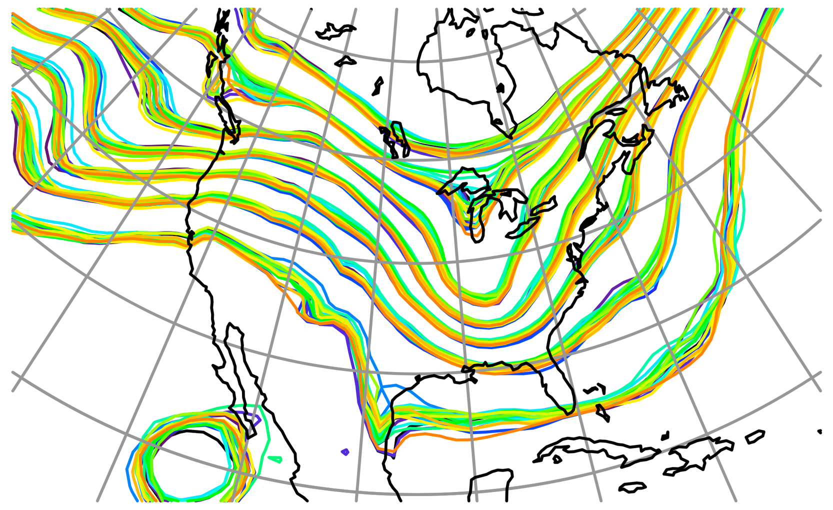

Empowering Geoscience with Improved Data Assimilation Using the Data Assimilation Research Testbed "Manhattan" Release

T. Hoar, K. Raeder, M. el Gharamti, S. Ha, A. P. Mizzi, N. Pedatella, J. Liu, Y. Zhang§,
C. M. Bitz§, H. L. Liu, C. Snyder, W. C. Skamarock, N. Collins, J. Hendricks, J. Anderson, H. Kershaw¶

National Center for Atmospheric Research Boulder, Colorado, USA
§ Atmospheric Sciences Department, U. Washington, ¶ Brown University

dart@ucar.edu

1. DART Manhattan Release Highlights

- ★ Able to handle much larger model states. This is needed for higher-resolution and/or strongly-coupled DA with multiple components. Distributing the model state across all tasks during the entire filter run means no single task must store the entire state at any time.
- ★ One-sided MPI communication allows tasks to request remote data items from other tasks without interrupting their execution or arranging which data items will be needed in advance.
- ★ Computing the forward operators for all ensemble members at the same time leads to code that vectorizes better.
- ★ Native netCDF support eliminates a conversion step that translated between netCDF model files and a DART binary format file. This also reduces the high-water mark for disk requirements.
- ★ Ensemble data can be read and distributed across all tasks on a variable-by-variable basis, reducing the maximum memory requirements.
- ★ Diagnostic state space files are now written in parallel with state-space restart files, resulting in faster I/O and lower memory requirements.
- ★ Support for externally-computed forward observation operators.
- ★ Support for per-observation-type localization radii.

2. DART is ...

The Data Assimilation Research Testbed (DART) is an open source community software facility for ensemble data assimilation developed at the National Center for Atmospheric Research (NCAR). DART works with a wide variety of climate and weather models and observations and has been free and publically available for more than 10 years. Building an interface between DART and a new model does not require an adjoint and generally requires no modifications to the model code. DART works with dozens of models of varying complexity, including (but not limited to):

- weather models, e.g. WRF, COAMPS, COSMO, MPAS Atmosphere,
- components of climate models, e.g. CAM, POP, CLM, WACCM, MPAS Ocean, MITgcm-Ocean, GCCOM, ROMS, JULES, FESOM, CICE5,
- atmospheric chemistry models, e.g. CAM-CHEM, WRF-CHEM,
- ionosphere/thermosphere models, e.g. TIEGCM, GITM, WACCMX
- low-order and simple research models.

DART assimilates a wide variety of observation types including:

- temperature, winds, moisture from NCEP, MADIS, and SSEC,
- total precipitable water, radar observations, radio occultation observations from GPS satellites,
- ocean temperature and salinity from the World Ocean Database,
- land observations such as snow cover fraction, ground water depth, tower fluxes, cosmic ray neutron intensity, microwave brightness temperature, and satellite brightness temperature.

DART provides both state-of-the-art ensemble data assimilation capabilities and an interactive educational platform to researchers and students.

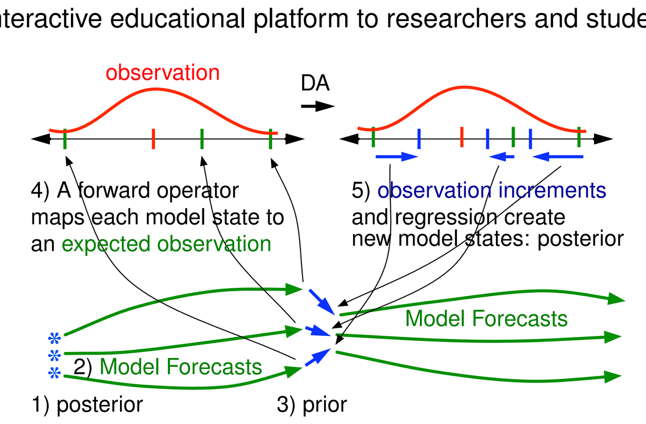

Figure 1: *Schematic for a toy ensemble size of 3.*

J. Anderson, T. Hoar, K. Raeder, H. Liu, N. Collins, R. Torn, and A. Arellano, 2009: *The Data Assimilation Research Testbed: A Community Data Assimilation Facility. BAMS* **90** No. 9 pp. 1283–1296

http://www.image.ucar.edu/DAReS/DART has information about how to download DART, the DART educational materials, and how to contact us.

3. Acknowledgments

The National Center for Atmospheric Research is sponsored by the National Science Foundation. Some computational resources were provided by the Computational and Information Systems Laboratory at NCAR.

4. Improved Adaptive Inflation

Author: el Gharamti

Ensemble data assimilation algorithms, especially with biased models, rely on algorithms that inflate the ensemble to avoid loss of spread. An enhanced adaptive algorithm that automatically determines appropriate spatially- and temporally-varying inflation has been developed, significantly improving on the previous method (‡ Anderson, 2009). The new method:

1. Models the difference between the ensemble mean forecast and the observed value as a random variable. This modifies the variance of the inflation likelihood by a factor of N_e^{-1} (ensemble size), such that the distribution is slightly shifted to larger distances.

2. Models the prior inflation as an inverse gamma (IG) distribution instead of Gaussian (N). This restricts the sampling of the inflation to positive values only.

The benefits of these enhancements are two-fold. The modified likelihood is expected to behave better in the case of a very sparse observational network and/or assimilating observations with large uncertainties. The use of an inverse gamma prior limits the magnitude of deflation, should it occur. Results for numerical weather prediction using a 2° Community Atmosphere Model (CAM4) assimilating wind and temperature observations from radiosondes, ACARS, AIRCRAFT, plus GPS radio occultation observations are shown below.

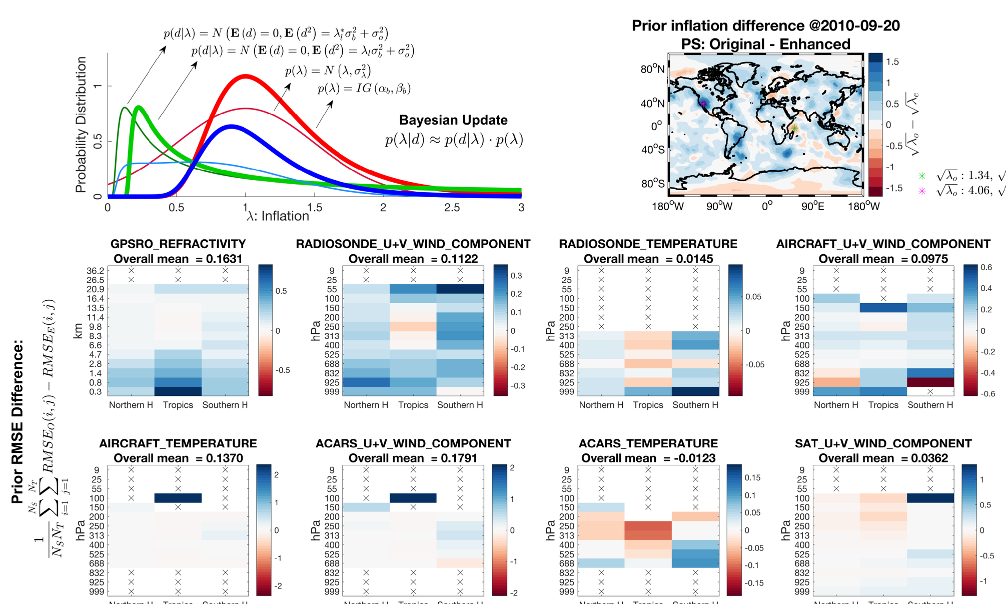

Figure 3: *Top left panel: Bayesian update of the inflation for a scalar case. The distributions (PDFs) refer to the prior, the likelihood and the posterior. Thick curves represent PDFs of the enhanced scheme. Top right panel: Difference between the original and the enhanced surface pressure inflation maps on September 20, 2010. Bottom 8 panels: Difference between original and enhanced, time-averaged, prior RMSE values as a function of model height and geographical region (blue means the new inflation scheme yields smaller RMSE). In the equations,* λ_l *= (enhanced) inflation parameter (* = old),* N_S *= number of observations,* N_T *= number of times,* $_o$ *= old inflation,* $_e$ *= enhanced inflation.*

‡ Anderson, J. L., 2009a: Spatially and temporally varying adaptive covariance inflation for ensemble filters. Tellus A, 61, 72-83.

5. Sudden Stratospheric Warming in WACCMX

Authors: Pedatella, H.L. Liu, J. Liu

The Whole Atmosphere Community Climate Model has been eXtended vertically to as high as ≈700 km (WACCMX). This enables the modeling of thermospheric and ionospheric phenomena, such as the ionosphere variability during sudden stratosphere warming events (SSW). WACCMX has been interfaced to DART, enabling the use of data assimilation to evaluate the performance of WACCMX in both model space and observation space. Fig. 4 is an example of the latter, for Total Electron Content (TEC) as measured by the Global Navigational Satellite System. TEC is not a model variable, and is not assimilated in these experiments, but can be calculated by WACCMX+DART for comparison against measured TEC. Assimilation of observations in the lower atmosphere improves the forcing of the upper atmosphere, leading to improved forecasts there (not shown).

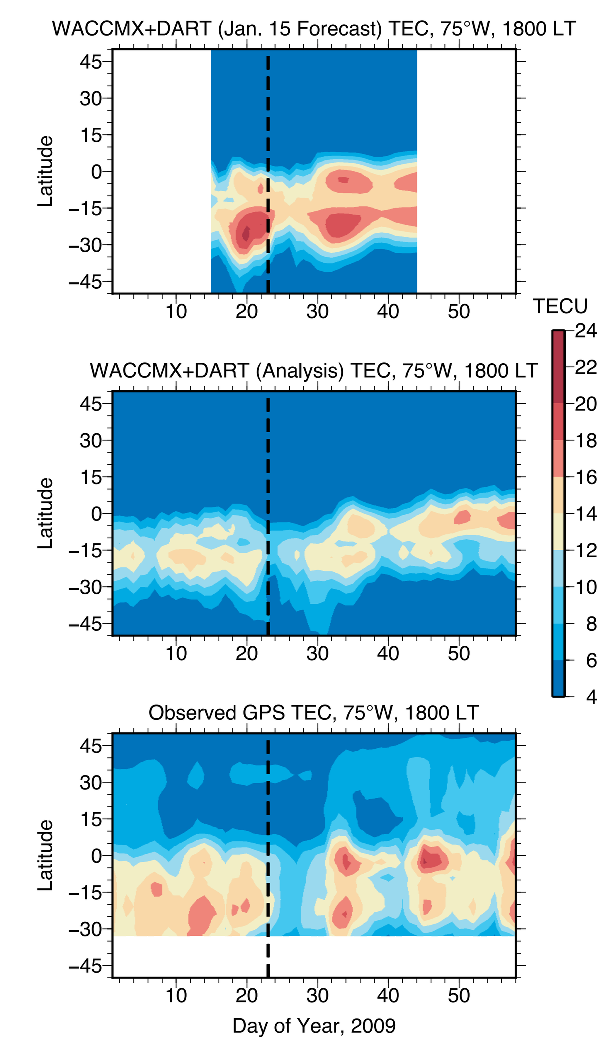

Figure 4: *The total electron content at 75W longitude and 1800 local time on successive days. Results shown are the WACCMX+DART forecast initialized on January 15 (top panel) from the WACCMX+DART analysis (ensemble mean) (middle panel), with ground-based GPS observations for comparison (bottom panel). The dashed lines mark the peak of the SSW.*

6. Sea Ice and Parameter Estimation

Authors: Zhang, Bitz

6.1 CICE5+DART: Real Observations

Retrievals of sea ice concentration (fraction of grid box covered by sea ice, SIC) from SSM/I microwave observations using the Bootstrap method were assimilated into CICE5+DART. The assimilation used adaptive inflation and and a localization half-width of 0.05 radians. CICE5 tends to overestimate sea ice along the ice edges and underestimate sea ice in the Central Arctic. In winter, DA can significantly remove the edge bias but does not influence the central Arctic as much. In summer, as the model ensemble becomes more uncertain in the central Arctic, DA is able to reduce the biases there too.

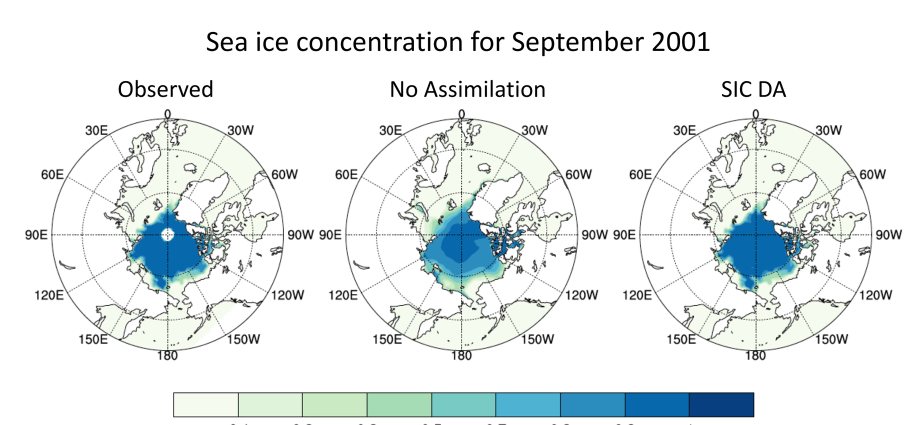

Figure 5: *Comparison of the September 2001 sea ice concentration (SIC) as observed by SSM/I (left map), from a free run of the CICE5 model (center map), and from a CICE5+DART assimilation using those same observations (right map).*

6.2 Parameter Estimation in a Perfect Model

Ensemble data assimilation can be used to generate estimates of model parameters using the 'augmented state' technique. Here the parameter R_snw, which contributes to snow albedo in the CICE5 model, is made part of the model state vector. Then the assimilation of any relevant observations can modify the value of R_snw to be more consistent with the observations. This is illustrated in Fig. 6 using a "perfect model" experiment, in which the true values of R_snw and the model state are known, but the assimilation is started with intentionally incorrect values. The assimilation guides the value of R_snw towards the true value. In parameter tuning experiments using real observations, in which the truth is less well known and the model is not perfect, the parameter may be guided to an unacceptable value. That value may allow better agreement between the model state and the observations, but violates the definition of the parameter. This points to the existence of model defects which should be fixed.

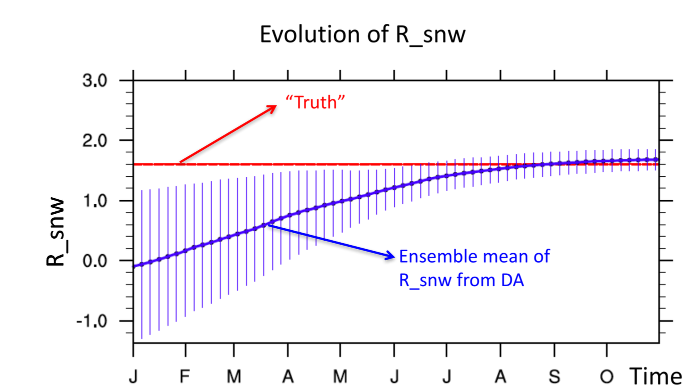

Figure 6: *In a "perfect model" assimilation, the distribution of values of model parameter R snw initially has the wrong mean (heavy blue line) and large spread among the ensemble members (vertical blue lines). Assimilation of observations taken from a free run of CICE5 guides the distribution towards the true value (red line) and reduces the uncertainty (spread).*

7. FRAPPE and WRF-Chem

Author: Mizzi

The Front Range Air Pollution and Photochemistry Experiment (FRAPPE) collected aircraft profile observations at six locations over the Colorado Front Range during nine special observing periods between July 20 and July 29, 2014. We applied WRF-Chem+DART to FRAPPE starting on July 14, with 6-hr cycling and 30 ensemble members.

We conducted two experiments:

1. a control experiment where we assimilated only conventional meteorological observations ("MET"),
2. a chemical data assimilation experiment where we assimilated MOPITT CO retrieval profiles and the meteorology observations ("RETR").

The left panel of Fig. 7 shows that RETR improved the 6-hr CO forecast throughout the troposphere by ≈10% compared to MET. The right panels show time series of the 6-hr forecast root mean square error (RMSE) and bias. For both figures RETR shows improved results compared to MET.

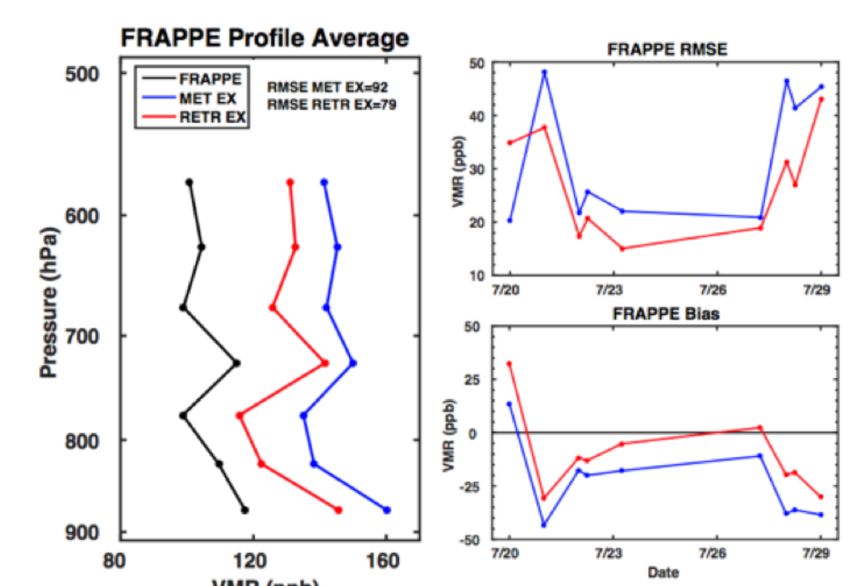

Figure 7: *Forecast skill improvements from applying WRF-Chem+DART to the FRAPPE. The black curve shows the FRAPPE aircraft CO profile observations. The blue curves show corresponding results from the control experiment where only conventional meteorology observations were assimilated. The red curves show results from the chemical data assimilation experiment where MOPITT CO retrieval profiles were assimilated together with the meteorological observations. The left panel shows the time and horizontal domain average CO profiles. The right panels show the domain root mean square error (RMSE) and bias of the assimilated states relative to the FRAPPE observations.*

8. Variable Resolution MPAS+DART

Authors: Ha, Skamarock, Snyder

The Model for Prediction Across Scales (MPAS) uses a global mesh (grid) whose horizontal resolution can vary continuously, minimizing the difficulties of nested grid models. To see the effects of locally enhanced resolution, we compare 5 day forecasts using a quasi-uniform global mesh with resolution of ≈120 km ("x1") and the same mesh with resolution increased to ≈30 km over CONUS (Fig. 8) ("x4"). Initial conditions for each forecast are generated by MPAS+DART ensemble data assimilation† on the native mesh. We assimilated all the conventional observations into a 96-member ensemble, which enables ensemble forecasts and robust conclusions.

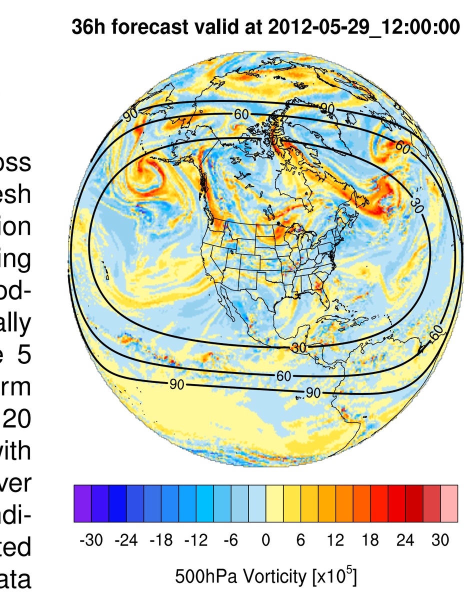

Figure 8: *Resolution of the 120-30 km mesh ("x4"), contouring every 30 km.*

8.1 5 Day Forecasts

The improvement in the forecasts due to using the variable resolution mesh in the initial condition analyses and/or the forecasts is shown in Figure 9. The forecasts are verified against the CONUS region of the NCEP final operational global 1° analyses (FNL) twice per day, but only on alternating days. The 5-day forecasts are improved more by using the refined mesh for both the initial conditions and the forecast, than for just the forecast.

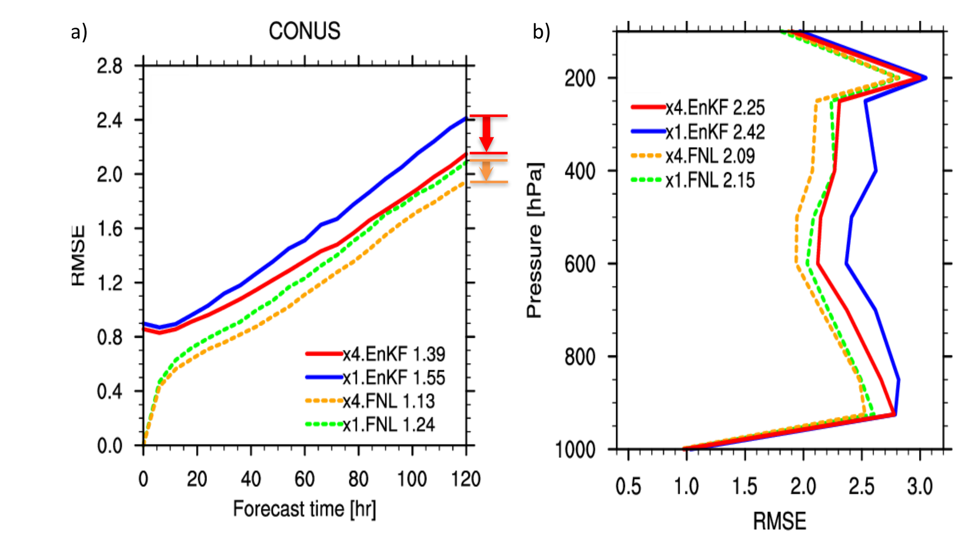

Figure 9: *When the NCEP FNL analyses are used for ICs (dashed), The refined mesh ("x4"), 5-day forecast, temperature RMSE is smaller than the uniform mesh ("x1"). When the forecasts start from refined mesh analyses (solid) the improvement is even larger (red arrow). This is true for the whole forecast period (a) and the entire troposphere (b).*

8.2 Power Spectra of Ensemble Analysis Increments

The ability of the refined mesh forecasts to generate smaller RMSE can be traced to the ability of their mesh to better represent the observations, which have no scale limitation. Figure 10 shows how far the assimilation process pushes the model state toward the observations for each spatial scale within the CONUS region. The analyses which use the refined mesh ("x4.EnKF") show much more increment power at spatial scales from ≈700 to ≈240 km because the uniform mesh model state cannot represent the observations as well as the refined mesh. Each curve is truncated due to its inability to resolve structures smaller than $2\Delta x$.

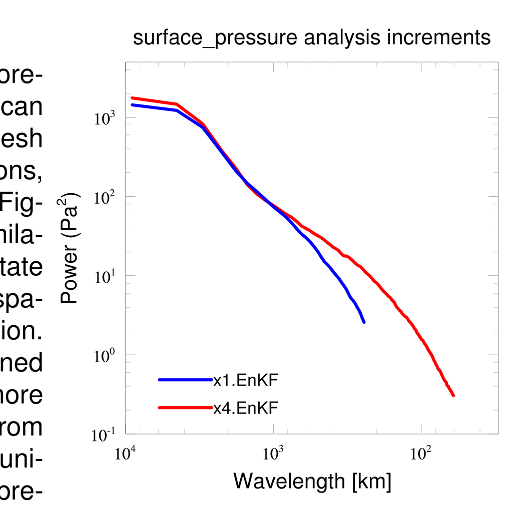

Figure 10: *Power spectra from assimilations using quasi-uniform and refined meshes.*

† Ha, et al. 2017: Ensemble Kalman Filter Data Assimilation for the Model for Prediction Across Scales (MPAS). Monthly Weather Review, https://doi.org/10.1175/MWR-D-17-0145.1.