

Challenges and some potential solutions for data assimilation

Discussion of the morning DA group

Presented by Xuguang Wang

CIMMS Predictability Workshop, University of Oklahoma, Norman, OK
Oct. 31-2, 2018

Background Error in DA

Challenges:

- model error and physics schemes/processes are insufficiently understood to be properly represented in background error.
- multi-physics ensembles can introduce new problems with clustering of members, introducing non-Gaussianity.
- forecast gains from the DA can be counteracted by biases in model components
- Advancement of treatment of sampling errors in ensemble based DA still needed

Solution ideas:

- stochastic physics with consistent model trajectory
- integrated development of DA and model physics (i.e. DA community work with model developers. see also “interdisciplinary” topic)
- may benefit from machine learning methods, especially for poorly constrained microphysics processes (see also Machine Learning topic)

Obs. error in DA

Challenges:

- need to better quantify observation errors including measurement errors, representative errors, observation operator errors, correlation of obs errors

Solution ideas:

- Collaboration between DA community and observation community (see interdisciplinary needs)

Computational bottlenecks for DA

Challenges:

- current operational DA code does not scale well
- I/O is a bottleneck especially with frequent cycles
- Preprocessing and QC of huge amount of data can be expensive
- latency of observations due to bandwidth limitations
- limited HPC resources, including both computation and storage

Solution ideas:

- “in core” DA can avoid slowdown from frequent I/O
- develop well-scaled DA algorithms and codes
- hardware investment for cpus and storage
- make use of cloud computing and/or cloud storage
- should all data be assimilated at all? better understanding and knowledge of observation impacts on forecast skills (see also Observation Impact topic)

Observation Impact

Challenges:

- under-studied for convective scales (especially systematic studies with many cases).
- complexity of observing system -- What are the interactions among different observation types?

Solution ideas:

- theoretical advancement using information theory beyond practical approach (OSE, adjoint, EFSO)
- well-designed experiments with agreed/appropriate verification metrics

Do we need new obs and have we fully utilized the existing observations?

New obs needed:

- obs that sample boundary layer: nation-wide 3D mesonet network consistently maintained with good quality
- direct obs/measurement to constrain microphysics
- obs to directly verify cloud and humidity forecasts
- observations relatively limited in mountain region and ocean regions.
- additional probes on aircraft could have value even if sub-research grade instruments.
- land surface observation networks

Do we need new obs and have we fully utilized the existing observations? (cont'd)

Under-utilized data:

- clear air radar Vr data
- dual pol data: need good microphysics schemes; efficient forward operators; unique QC considerations. Assimilate dual-pol derived products vs direct assimilation of dual pol signal.
- non traditional networks: cell phones, truck data, etc.
- visible imagery: very expensive for RTM because of 3D scattering but may have value for very high resolution.

Other challenges and needs:

- limited access to international data (e.g. radar from other countries). lack of sufficient protocols and data sharing agreements.
- unified data repository with qc, uncertainty, meta conventions and with proper documentation. Utilize cloud storage.

DA algorithms

Challenges/needs:

- need to investigate methods to deal with and measure imbalance between analysis and model for convective scales
- How to assess the impact of the advancement of different component of the DA algorithm (background error vs ob error vs nonlinearity etc) to prioritize efforts?
- Methods to treat nonlinearity/non-gaussianity. Should we move away from current EnKF/EnVar framework to pursue non-parametric or parametric non-gaussian filters? How to quantify the impact of violating nonlinearity/gaussianity assumptions on analysis and forecast?
- How well would 4DVar with tangent linear of simplified physics work for convective scales?
- observation targeting theory is limited in the context of nonlinear convective scales

Fine resolution DA

Challenges:

- require rebuilding of observation operators and physics parameterizations
- need to deal with nonlinearity and non-Gaussianity
- need to deal with correlated observation errors given less thinning/superobbing
- computationally expensive
- More understanding of the impacts of fine (sub-km) resolution DA on different processes

Can machine learning be leveraged to advance DA?

- Need to better understand ML algorithm and its capabilities and limitations in collaboration with ML experts
- Potential applications in DA e.g.
 - ✓ soil state estimation
 - ✓ preprocessing observations, QC
 - ✓ DA parameters
 - ✓ Replace TLA

Multi-scale/coupled DA

Challenges:

- Require models to better simulate coupled processes
- Develop DA algorithms to effectively extract information from all observations to update/correct system that houses various scales both in time and space.
- Next generation DA and NWP: should we have a global CAM with rapid update to fulfill all applications of all scales or continue using a nested approach?
- Better understanding of relative benefit/impact of updating e.g. small/large scales

All sky satellite radiance DA

- Improve model to better represent cloud process
- Improve CRTM especially for ice phase
- Need consistent assumptions in CRTM lookup table and microphysics schemes in numerical models
- Bias correction:
 - ✓ separating sources of bias (ob vs model vs RTM [esp. in ice phase clouds]) – need good independent observations
 - ✓ separating bias for clear air and cloudy radiance
- Using satellite radiance to identify model errors
- retrievals vs radiances?

multi-disciplinary needs to advance DA

- create environment to promote multi-disciplinary solutions or integrated development
- ✓ Collaboration between DA community and model physics developers to (i) better present model errors in DA and to (ii) use DA to better identify/quantify/understand model errors
- ✓ Collaboration between DA community and observation community on (1) operator development (2) obs error estimation (3) observation network/sampling strategy design
- ✓ Leveraging machine learning during the end to end process of DA requires collaboration with machine learning community