Working group recommendations -- DET Responses
Created September 14, 2010
Modified February 24, 2010

Overall design
(Det1.1) <u>Question:</u> Do you see a need for DET to have real-time capability? If so, where?

<u>Recommendations:</u>
[WG 1 Session 1] Yes. Already have an ensemble system (see above). As long as there is resources (CPU and disk space). Leverage efforts from HMT and HWT. Create a data set for evaluating modules. Subjective forecaster evaluation.
[WG 2 Session 1]
- Long-term goal should be to emulate NCEP's system as closely as possible.
- Group splits on priority level of in-house real-time capability; difficult to make concrete recommendations now.
    - Benefits include forecaster and user feedback and fault tolerance/system hardening.
    - Drawback is complexity and cost.
- Strongly encourage outreach to/collaboration with University groups who are running in real time to focus on DET application testing-benefits forecaster and user interaction/feedback and ingestion of the University capability
[WG 3 Session 1] While this is not a requirement for DET to serve its essential functions, near real-time would be valuable. For instance, if the near real-time capability already exists or could be added without substantial cost, it would be useful.

<u>DET Response:</u>
Based on these recommendations, DET will develop their system for both real-time and retrospective capability with initial focus on real-time and allocation for retrospective throttling in the future.

(Det1.2) <u>Question:</u> How do we make sure tech transfer to operations is possible?

<u>Recommendations:</u>
[WG 1 Session 1] Needs to be useful for different operational groups. Different users have different needs (e.g., Air dispersion – PBL, military, DTC…). Needs. DET person at NCEP collaborating (currently: GSI person, monthly teleconferencing, etc…). Staff exchanges between DTC and NCEP. Testbeds have other users that can contribute (how can we use it). Can help with planning and selling the software.
Other questions: How long to test to make operational? At what point do we decide if ready for operations? Follow close as possible to EMC to go smoothly to operations?
[WG 2 Session 1]
- Encourage compiler continuity, standard coding practices.
- Computing platform differences/ages, making direct transfer more difficult, recommendations: less.
- DET could be viewed as testing for pending/future computational capability.
[WG 3 Session 1] Comfortable with current plans, keep working closely together.

<u>DET Response:</u>
Based on these recommendations, we will continue to plan for compiler continuity, standard coding practices, standard data formats, etc.. We will also plan to continue coordinating with EMC for the

transition to operations. There are also plans to have staff exchanges between DTC and NCEP, including operationally useful testing platform.

(Det1.3) <u>Questions:</u> How do we identify a list of candidate methods to consider for testing?

<u>Recommendations:</u>
[WG2 Session 2] Ask what other centers (UKMet, CMA, ECMWF) and Universities are doing; Set "entrance criteria"; Standardized skill level for specific parameters, related to performance of the benchmark ensemble for specific domain (e.g., CONUS), period (e.g., all of 2009), etc. ; Make sure Implementable

<u>DET Response:</u>
B

(Det1.4) <u>Question:</u> How do we insure plug-and-play and portability?

<u>Recommendations:</u>
[WG3 Session 1] Keep dependencies to a minimum. Be involved in other initiatives working to make common software structures (i.e. ESMF). Ensure all input data formats are accounted for. Have well documented test data and methodologies for creating new products, both from DET and community.
<u>DET Response:</u>
B

Module 1: Ensemble configuration

(Mod1.1) <u>Question:</u> What other system should we consider in our initial investigation? Who are the contacts for the systems?

<u>Recommendations:</u>
[WG 1 Session 1] Want scripts to run various components easily at DET (others out there: DART, CAPS system,…). Need more DART modules-Example: Ryan Torn (You.Albany-runs DART ENKF).Other systems-Stony Brook University--Colle; Navy?
[WG1 Session 2] Jim Hansen – DART; Teddy Holt – NAVY (COAMPS); Bill Gallus – ISU; Ken Mylne – MOGREPS; COSMOS (Europe); CMC (Canada)

<u>DET Response:</u>
We will add these systems to our list.

(Mod1.2) <u>Question:</u> What are the similarities and differences between how the current ensemble systems are engineered?

<u>Recommendations:</u>
[WG 1 Session 1] Issues:
- One executable or many?
  - One executable-allow ensemble members to interact.
- Is it ESMF?

- Fault tolerance.
- Scripting language.

DET Response:
We will keep these issues in mind while developing the system. EMC tends to prefer many executables to one, NEMS is an instance of ESMF, our real-time system will require fault tolerance; our scripts will be written in commonly used languages.

(Mod1.3) Question: Does the community (research and operations) feel that they might want to adopt/ use the DET Ensemble Configuration and Initial Perturbations modules? If so, what will it take to facilitate this?

Recommendations:
[WG 1 Session 2] University community = yes or no, depending upon robustness of existing system at a University, and complexity of the DET system. The more experienced the user, the less likely they are to use a DET system. It is possible that users will select portions of DET tools to use. There are benefits of using a tool that has been vetted by the community.
Operations community = yes, but slowly. AFWA usually imports whatever the research community produces. AFWA usually has a lengthy testing period before becoming operational.
University Issues: a lot of effort to learn the system. A streamlined, simple system, is more likely to be used (shorter spin up time learning curve).
Operational issues: is it designed to meet our needs (moving windows, limited obs regions)

DET Response:
We

(Mod1.4) Question: Other comments or questions?

Recommendations:
[WG1 Session 2] How many ensemble members do we need? 80 versus 20 versus 10 versus 2. Issues of a single executable (with 80 members, the odds are higher of one of the members failing and bringing down the entire workflow).

DET Response:
W

Module 2: Initial perturbations

(Mod2.1) Question: How should the infrastructure be organized to accommodate testing of various IC and LBC perturbation approaches?
Recommendations:
[WG 1 Session 1] Interpretation-How general sure we make the system?
- First, system must run ensembles in DET
- IC and LBC in one format (similar to NCEP – EMC and WRF: . Ornette CDF)-do not want multiple formats on disk and for archiving.
- User required to put into a common format.
- Look at DART for formats-ensembles in the data assimilation context. DART-WRF--ask C.

Snyder space what is preferred?
- Depends on what you want to do… Borrow from different systems.
- WRF GUI portal to run ensembles and retrospective mode (GSD).-Starting point?
- CAPS multi-model –> space converters to one output format for different models (NMM, ARW, ARPS,…)

DET Response:
We recognize the need to aim for one format for both IC and LBC. Based on our experience with MET developments at DTC, pushing for one formats may be slightly unrealistic and may cause the community to not be able to use our tool. One solution may be to provide a suite of preprocessing tools that will output a common format. Discussions during the workshop regarding looking at dart for formats suggested that dart does not have a specific ensemble format that we should be considering. WRF GUI portal and other configuration modules for WRF ARW and ESMF will be considered. DET is developing a strong working relationship with CAPS and we anticipate benefiting from the work they have done to converter model output to a consistent format. A single format most likely also be addressed by the release of the Unified Post-Processor code by DTC early next spring. This code is geared towards reading in and writing out many of the common data formats currently in use.

(Mod2.2) Question: DET is considering collaborating with GSI task for ICs space and for growing perturbations. Is this a good approach?

Recommendations:
[WG 1 Session 1]
- Difficult to answer at this time. Probably good idea. Treating ensembles and data assimilation separate is probably not a good idea. Need to think of a coupled system.
- How can you test different ICs without including data assimilation? Coupled issue, so cannot ignore. Data assimilation part of the system.
- GSI-need to support what is going at NCEP-transfer to operations.
- Use GSI to enhance other ensemble systems (ENKF).
- DONE: DET-GSI-focus on research to operations-repository for code for community.
- Look into NCAR Data Assimilation Team for additional ideas on data assimilation issues.
- DET collaborating with research community. GSI-hybrid.
- How to do this collaboration?
- Ensemble enhance component to the GSI? Not decided…

DET Response:
We recognize that treating ensembles and data assimilation separate is not the best approach. That is why we propose to work with the GSI task and the NCAR/RAL Data Assimilation Team to identify which approaches are pertinent to operations. We do however feel there. There are 2 separate dimensions to this problem and that we can change one without losing the coupling. We will continue to keep the community apprised of the outcome of these collaborations.

(Mod2.4) Question: What options for IC and LBC perturbations would the community suggest we focus on first for testing?
Recommendations:
[WG 1 Session 2]
- Use existing perturbations (baseline = NCEP SREF)
- Use something that does not involve data
- Something more ensemble based

- LBC =   For cold starts, the best and most simple route is an existing operational ensembles (global US and/or CMC, or SREF).
- IC =   Existing SREF/GFS perturbations
- ETR method or cycling method, such that it is consistent with the boundary perturbations.
- EnKF has two steps, the analysis and the perturbations, perhaps the two steps can be split? EnKF is ready for testing, whereas the hybrid is still a little ways away. (Issues about multi-core systems)
- DART (incorporating Ryan Torn's scripts for managing the forecast)

DET Response:
Noted.

(Mod2.5) <u>Question:</u> Other comments or questions?

Recommendations:
[WG1 Session 2] What does the regional ensemble need to do – for example, regional ensemble likely does not need to assimilate satellite winds if that ensemble uses GFS ensemble, as that data is already incorporated.
Need to rapidly deploy to data sparse areas, a nod to more simple schemes.

DET Response:
Noted.

Module 3: Model perturbations

(Mod3.1) <u>Question:</u> How can we have a stochastic forcing that is more generalized for use with several model cores?

Recommendations:
[WG 2 Session 1]
- Definitions:
  - Tendency forcing
  - Internal physics forcing
- Tendency forcing
  - General tool set/modules that could be plugged into different models to draw spatially and temporally correlated structures, subject to amplitude determination and specific model variables.
  - Module to apply multi-scale (forced backscatter) perturbations.
  - Going much more general is difficult (if not impossible)
  - DET/DTC has a role to make tuning of these schemes more efficient.
- Internal physics forcing
  - Pre-tuned/tested parameter distributions could be allowed for models that already have the code built in (e.g. Grell Cu scheme).
  - Scripting layer could support namelist options for those models that do support e.g. parameter values or stochasticity.

DET Response:

(Mod3.2) Question: What methods and metrics should be used to separate the uncertainty generated by initial perturbations versus model core uncertainty?
Recommendations:
[WG 2 Session 1]
Methods are not known well enough. DET/DTC supports this research question by facilitating easy ensemble runs using both model and IC uncertainty (and DA).
DET Response:

(Mod3.3) Question: Are there suggestions for how to execute core interchangability (i.e. ARW, NMM-e, NMM-b initially but others in the future)?
Recommendations:
[WG 2 Session 1]
- Specify requirements for input and output.
- DET provides regridding tools for input and output (NEMS could help).
- Use small amount of DET/DTC resources to collaborate on working out interface details (consulting, etc). Maybe visitor program to facilitate.
- DET should encourage common map projections being adopted by any participating models.

DET Response:

(Mod3.4) Question: Is there clear guidance on which to test first: Single model with stochastic forcing or multi-model?
Recommendations:
[WG2 Session 2] SREF 2011 (multi-model) is the benchmark; No clear guidance on testing, except National Mesoscale Probabilistic Prediction White Paper recommends research thereof; Research improving/optimizing multi-model ensemble, then compare to performance of single model ensemble with thorough stochastic forcing; Test combining both approaches
DET Response:

(Mod3.5) Question: Other comments or questions?
Recommendations:
[WG2 Session 2] Should DET use 1-D model to cheaply investigate some research questions?
- (e.g., investigate sensitivities of stochastic forcing techniques)

DET Response:

Module 4: Statistical post-processing

(Mod4.1) Question: What test data should we consider for Statistical Post-Processing?
Recommendations:
[WG2 Session 2] Need lots (years?) of data and observations at many scales
- Training (dependent) dataset – 2008?
- Validation (independent) dataset – 2009?

- "ground truth" data – gridded analysis and observations
  - Need information on error in the ground truth

Module 5: Products and display

(Mod5.1) Question: How do we interface with AWIPS II development.
Recommendations:
[WG3 Session 1] Need more information on when AWIPS II will be deployed and what the follow-on development will be. DET should not expend resources, but should only play an advocacy/awareness role. Look beyond to NexGen.
DET Response:

(Mod5.2) Question: Should we think beyond AWIPS II to web-based displays for the research community?
Recommendations:
[WG3 Session 1] Web-based displays should be pursued. Web-based is primary for AFWA, also useful for testbeds.
DET Response:

(Mod5.3) Question: How should we include social scientists in this portion of the testbed?
Recommendations:
[WG3 Session 1] Engage social scientists to participate as possible in testbeds and to also pursue external funding sources for more extensive participation.
DET Response:
W

(Mod5.4) Question: What is a prioritized list of ensemble-relevant forecast products?
Recommendations:
[WG3 Session 2]
- Ensembles absolutely need to be researched and developed so that they become a useful tool for alerting users to high impact events
- If ensembles can provide useful info on impact high events, how do you extract it and display it? Will need to meet various levels of complexity
- What kind of forecast info is required?
- Garden variety weather impacts many users (e.g. energy use, financial market, agriculture)
- Tools for developers, researchers
DET Response:
We will explore these products.

(Mod5.5) Question: What is a prioritized list of ensemble-relevant display algorithms?
Recommendations:
[WG3 Session 2]
- How people are going to display data
- Interactive tools and displays that user controls
- Spaghetti-grams, make more useful
- Sounding diagnostics for ensembles
- Effective visualizations that "grab" attention
- Web II to address interrogation needs

- Mobile technologies ("iP" whatevers)
- ENSEMBLE MEAN: Not on model or atmosphere's attractor

DET Response:
We will explore these display algorithms.

(Mod5.6) Question: How do we develop and solicit new display possibilities? (Ensemble Products Workshop II?)
Recommendations:
[WG3 Session 2]
- Ensemble products workshop-YES
- Strong representation of appropriate end users and decision makers at workshop
- AMS presentation?
- Search for "entities" that have *avant garde* tools and ideas that address user needs
- Start a forum to search info

DET Response:
We will explore these avenues.

Module 6: Verification

(Mod6.1) Question: Do we need to be discussing the inclusion of MET (or MET-like capability) in AWIPS II?

Recommendations:
[WG3 Session 1] DET should adopt an advocacy role as well as look forward to NexGen.

DET Response:
Based on this recommendation, we will consider NexGen requirement more closely.

(Mod6.2) Question: What is a prioritized list of ensemble-relevant verification products?

Recommendations:
[WG3 Session 2]
- Basic tools will be there
- Time-space scales, user, phenomenon dependent
- Account for various needs of all users
- Identify issues that ensemble products, prob fcsts will provide values
- Appropriate measures for high impact weather
  - What constitutes high impact and/or rare event?
- Point (rich history) vs. object verification (new)?
  - Spatial verification not mature for ensembles
- Decomposition tools

DET Response:
We will explore these verification product ideas.

(Mod6.3) Question: How should we filter through ideas provided by other workshops (i.e., RAL Verification Workshop 11/2010; WMO Verification tutorials; others)?

Recommendations:
[WG3 Session 2]
- Focus sessions for ensemble verification at workshops?
- "Literature review" by aware individual

DET Response:

(Mod6.4) Question: What analysis fields and/or observation data should we consider for verification?

Recommendations:
[WG3 Session 2]
- Both!
- Incorporate uncertainties assoc. with fields/data
- Use ensemble analyses (e.g. Torn and Hakim)
- Analyses OK for planetary <-> synoptic scales
- But they are not as valuable for smaller scales
- Sensitivity of verification against various analyses
    - QPF/PQPF examples; others?
- Ceiling analogy; how do you fill the gaps?
- Everything needs uncertainty disclaimer

DET Response:
We will explore these display ideas.

Testing and Evaluation Protocol
(TE1.1) Question: How much rigorous end-to-end testing is required vs. evaluation of individual components?

Recommendations:
[WG 1 Session 2]
- Test components twice, end-to-end once.
- Robustness of skill in many situations: (tropical, mesoscale, synoptic scale, boundary layer, aviation/flight level).
- Robustness of system (99% completion, versus 90%). Failure tolerance. Track the failure rate of end-to-end system.
- Track flop counts, message passing. Information helpful to hardware design.
- Cost/benefit analysis: how much is needed (CPU, disk storage) to generate forecast improvement. More but coarser ensemble members – more lets you do EnKF, but perhaps not resolve convective scales.
- Full documentation of testing of different approaches, for different parameters. The added value of the different approaches. Let DET be the broker of the strengths/weaknesses of the different methods.

[WG2 Session 2]
- Should follow EMC guidelines/standards
- NCEP bundles implementations 1/year
- min of 30 days, 4 seasons

DET Response:

(TE1.2) <u>Question:</u> What different metrics would you suggest for different evaluation plans? (please give evaluation goal and metric)

<u>Recommendations:</u>
[WG3 Session 1] The metrics should be appropriate for the test plan.
- **What different metrics would you suggest for different evaluation plans? (please give evaluation goal and metric)** *(repeated in case others have joined the group)*
  - **Depends on time, space, phenomenon, user**

<u>DET Response:</u>

Collaborations
(Col1.1) <u>Question:</u> How can DET feedback into NUOPC and the Probabilistic Prediction Advancement Committee (PPAC), including the PPAC Business Case?

[WG 1 Session 2] Invite NUOPC to the DET meetings and vice versa, have a DET 'delegate' for NUOPC meetings. Incorporate common model architecture committee. Incorporate content standards committee. Incorporate TTP Metrics committee. Let PPAC leverage the DET community to promote their agenda. AFWA has experience in tapping into the work of the community, as does the private sector. Have the DET go to NUOPC and PPAC to understand their needs/requirements/requests.
[WG2 Session 2] Common people with NUOPC; Design PPAC working groups to be intimately tied to DET; Mold PPAC WG's to match up with DET modules (e.g., "Business Case WG" and "Interface WG" become the "DET Products and Services Module WG"); Establish WRF Ensemble WG as one of the PPAC WG's; Include DET personnel on PPAC WG's
[WG3 Session 2] Building case for ensemble advantage; Provide relevant cost-benefit analyses to potential stakeholders; Entice businesses to join ensemble movement

<u>DET Response:</u>