Blue Group Notes:

Reporter: ?

Overall question at start**:**

Cliff Mass: why DA/ENS? Isn't it much bigger than this?
A: issues bigger than that, they want this community's input

1) What have you heard about EQ that most excites you

Rob Fovell: Example, WRF was not one group project, became a resource that can be leveraged. I like the idea of a coordinated national ensemble. One of his students working on model physics; he is going to need to make 100s/1000s of simulations … but, if there is a common resource/database, it would facilitate these kind of forward advances without duplication.

Brian Jewett – potential unexpected opportunities to work with other communities..

Nick: increasing diversity, getting more women in field; combine with classes that could perhaps make runs in the classroom.

Kevin: excitement about ensemble: our grad students and ugrads most want to run a model. What I'm hearing is there is more that can be done. Emphasis now on verification and validation, taking that to faculty and students will help.

Glen: potential for making it easier to access data particularly observations, and to do verification and find out what you want to learn about the model without saving history files. We compute a lot more than we save; now being told we can't save everything anymore. I'm hoping there will be things coming out in both ingest and output side that will help, front-end and back-end, that need help right now.

On cyber-side: ability to process data on server side, everyone recognizes how important that is. Open source science idea is really promising, collaboration in larger space.

Main concept for EQ is getter out profession looking at longer time horizon. He is less excited at each university running a version of wrf. How are we going to help learn data discovery etc; that is what is really lacking.

Discovery-to-knowledge… NSF not used to working with lots of data, unlike nasa etc. Need interagency work to improve methods –

Like the idea of having data accessible from one location, maybe a wiki for data, where can talk about problems with data or quality control.

Mark Stoelinga: turning 80/20 into 20/80. Envisions spring thaw in which graduate students can do both of these things now.

Chris Snyder: potential for NSF to fund infrastructure that would make it easier to collaborate as a field. I'm not sure what format that would take. We have to get behind it somehow to get NSF to do the right thing.

Cliff Mass: we work too independently and in isolation, and to him interactive capabilities are the easy thing. Key thing is we don't work together.

Brian Colle: encouraged to hear there are testbed-type of activities, and education (Rich) – so some seeding efforts are underway we can learn from. It did seem to me that a lot of these issues are local; I do wonder how much centralization [is practical]. Tough for me to imagine a cube that will make everyone happy; there are local challenges.

Most exciting is university-based national ensemble. I see the community becoming aware of the data. Disappointment – so many tools – [spread in ensembles]

Use of satellite data – if have structure and community assimilation model; we can do sensitivity studies … what broad applications.

2) What key challenges keeping us from going forward in advancing DA/ENS
3) Science drivers in DA/ENS

Russ: big driver is improving probabilistic prediction; small is improving individual schemes.

- main obstacle is computer. #2 is not organized.

Chris Snyder – how we maintain large systems (software; but could be large community considerations) – we can really use some good ideas, perhaps from computer science; on science side, once you have a big model, how do you make it better?

Response: there are difference governance models out there; Apache foundation is one model.

Obvious science drive on DA side is how to specify the current state of atmosphere. DA designed to improve a 48-72 hour forecast are not necessarily applicable for other uses.

4) Infrastructure advances

utilize data management people, worked with CS people who were given a TB of data, who came back with knowledge of seasonal and daily cycles…
chris: predictive models provide a lot of information, we work differently…
Russ: I don't know lots of CS people at my department…
: Penn State has data specialists, they are looking for applications

5) Central vs. decentral coordination

We don't want to have one group to run everything. As we build more central structure, there are governing rules that will limit us. Independent/new ideas will be brought up. Viz: there isn't just one.

Need more input from library and information science. John's point, need a wiki of information. Some group is needed to be vested with making a catalog of this data.

Russ – has page on which people can leave questions –

Rob: what if you asked Wolfram Alpha, what is windiest day …

What if every time you asked for a data set, you also got a citation.
Other data centers already supply citations.

Russ: are same efforts for Cliff to improve NW forecasts different from Brian's for NE?

Cliff: how do you measure quality of a forecast? Testbed can help you –
Russ: can we even agree on measures?
Cliff: USAF can. DTC does this with them all the time.
Me: could use societal impact as one measure, precipitation, wind speeds…
Nic: could get right answer for wrong reason.
Cliff: could do targeted research; could have compensating errors but need to do it.

Chris Snyder: hip-bone is connected to the thigh bone. As soon as you have a continuing cycle, run over a larger domain, you will see these errors show up. Contrast – if run model off someone else's analysis, model errors could compensate for analysis errors etc.

Brian Colle: would rather spend energy and coordination in targeting problems sooner rather than later. There is a communication issue –
Russ: there are things you might have identified in the models that you haven't shared that I might encounter … having this information is crucial
Brian Colle: we've had workshops and prepared reports…
Russ: would be nice if when downloaded WRF you had some information right there about what physics work well together…
Rob Fovell: you've mentioned workshops I wasn't even aware of so I had no benefit from it…

Chris Snyder: should be able to publish a result of e.g. a month's simulation and what you encountered.

Nic: was BAMS papers 2 years ago/2010, here is how to conduct parameterization research.
Chris: but this is the big data problem.
Me: this was to limit publications on this type of study –
Chris – it does no harm, I would love to be able to search for that…
Russ: this is partly what DTC is supposed to do with their reference configurations.

6) Morgan's question: what is broader impact of EQ in future
related to? – (7)
7) What are concrete next steps

Glen: need to bring others, including statisticians in, those much better at that than we. Same with CS specialists…
Cliff: can't just throw people together. Need sustained interactions with these people – have to learn each other's language before [progress]
: In MD, lot of co-advising going on, long time periods

• facilitate meaningful, long-term collaborations between interested disciplines e.g. atmos sci, CS, statistics, math. Reduce "discipline friction", lead to culture change.

Ethan: EQ started more in cyberinfrastructure; need to bring in domain experts.

• coordination between model and verification to diagnose and understand model errors (and the processes that cause them), how to know if we're making progress, how to best do probabilistic prediction and how to know whether we're making progress on that

 Glenn: how to work with 'dirty' data.
(are we making progress? what are our metrics?)
Question of: which physics packages are good …
Model with worst bias could provide best prediction, if readily correctable.
How do we do this for ensemble prediction and DA?

How to put this into practice?
None of this is hard, technically.
Chris: easy to diagnose model is wrong. Why is hard.
Russ: if I have resources to do what I want to do – why work with someone else.

• Being able to get all the same observations in a consistent usable format with information about quality, etc.

• <mark>Online electronic journal to put information out there regarding all of the knowledge about model errors and biases and results that others have found</mark>

Others can create this using common tools … important to be able to search it (I referenced ACM computing classification system). Russ: JAMES. http://www.agu.org/journals/ms/. Rob: electronic journals have in general not been too successful.

8) If there are people or groups, disciplines that we missed –
not addressed