Four types of dynamical downscaling:

1. Limited area model (LAM) forced by lateral boundary conditions (LBCs) from a NWP GCM or reanalysis at regular intervals (typically 6 or 12h), by bottom boundary conditions (e.g., terrain, SST) and specified initial conditions (ICs). This type of downscaling is in weather prediction mode so that the ICs are not yet forgotten. Predictability: Day-to-day weather prediction.
2. LAM initial atmospheric conditions have been forgotten, but results are still dependent on the LBCs from an NWP GCM or reanalysis on the bottom boundary conditions. This type of downscaling is in regional climate simulation mode. Predictability: Seasonal weather simulation.
3. LBCs are provided from a GCM which is forced with specified surface boundary conditions. (e.g., Global model forced by observed SST). Predictability: Season weather prediction.
4. LBCs from a completely coupled earth system global climate model in which the atmosphere-ocean-biosphere and cryosphere are interactive. (e.g., IPCC scenario; US National Assessment). Predictability: Multiyear climate prediction.

Type 1 is in weather prediction mode while Types 2 through 4 can be considered RCM modes. Observed data does not exist in Type 4 simulations.

Discussion 1: The values added of dynamical downscaling.

1. **Type 1 downscaling:**
   - We all agree that this type downscaling add values.
   - Questions rose by Yongkang:
     1. In this type of experiment, initial conditions (ICs) are dominant. What about LBCs? Does GCM outputs could be used for LBC?
     2. Although RCM as a group has the ability for this type of downscaling. Does every RCM have the ability to downscale properly for every region in the world? His feeling is not. As Barry pointed out, physical processes play a role. So the dominant physical processes in a specific region, such as convective processes, moisture transport processes, and/or land processes, have to be properly represented in a specific RCM for this region.
     3. The successful downscaling is due to the memory of initial conditions. How long this dominant effect will last? It seems we have no question for initializing every 24 hours then looking at long term mean for climate (as did in Xue et al., 2001, Mon. Wea. Rev.)
   - Points rose by Barry:
     1. Although ICs are dominant in this type of downscaling: a) physical process b) position of the domain boundary and c) quality of the LBCs do play a role in the downscaling.
   - Points rose by Roger:
     1. The downscaling from the weather prediction models is a Type 1 downscaling, and, he agrees, does add significant predictive skill (i.e.

value). The reason is that with Type 1 downscaling (say from the GFS), the regional model runs (e.g. from WRF) include observed initial conditions within its domain.

2. **Type 2 through Type 4 downscalings:**
   - No accordant conclusions are made on whether these types of downscaling add values.
   - Nudging would help on Type 2 downscaling if forcing fields and LBCs are from reanalysis.

   - Points rose by Roger:
     1. Initial condition knowledge is lost in Type 2, 3 and 4 downscaling. The regional model only has the lateral and bottom boundary conditions to insert information. With Type 2, the lateral boundary condition information is based on observations (i.e. reanalyses), yet we show that the larger scale (synoptic information) deteriorates unless nudging, particularly spectral nudging, is used. With this nudging, however, the regional model is now significantly constrained by the larger scale reanalysis. When you apply spectral nudging to the Type 3 and Type 4 simulations, the regional model, at the synoptic scale (i.e. the scale resolved by the larger models) is constrained by the larger model (with all of its biases).
     2. RCMs cannot skillfully create the synoptic structures that we need. He uses the term "value-added" for the synoptic scale to mean that the accuracy (fidelity) of the simulation of features that are resolved by the larger scale model as improved using the regional model. For Types 2, 3, and 4, this does not, unfortunately, appear to be the case.
     3. Spatial structure is added in these types of downscaling, but since it is so dependent on the larger scale model, biases in the input from that model cannot be corrected by the regional model. This is shown convincingly to me in Lo et al. 2007, since the interior results are so dependent on how frequently the regional model is reinitialized.
     4. Thought experiment: Assume the region we are interested in is over open ocean with no surface SST gradients. Can the regional model provide increased accurate spatial and temporal structure?
        - For this experiment, it means the added information that the model has for Type 1 downscaling are the initial and lateral boundary conditions. For this case, there is a lot of evidence that the regional (weather) model adds skill at both the small and larger scale features without nudging of any sort even over the open ocean.
        - For Type 2 downscaling, the reanalysis (which provides a relatively coarse sampling of continuous atmospheric fields) provides the only added information. If nudging is used, this real world data can be used to constrain the larger scale features, and let the finer scale features evolve as accurately as possible given the regional model dynamics and parameterized physics. Of course, the regional model than

becomes dependent on the reanalysis for the larger spatial features even within the regional model domain.

- For Type 3 downscaling, a GCM is run but with prescribed surface forcings (e.g. SSTs), so the constraint by the real world is much less than in Type 2 downscaling.
- For Type 4 downscaling, all real world constraints are removed except, for example, the atmospheric well-mixed greenhouse gas concentrations. Even if nudging is not used, the regional model is still dependent on the GCM for the larger scale features through the lateral boundary conditions, and, thus, in that sense is a slave to it. Biases/errors in the lateral boundary condition values cannot be corrected by the regional model.
- The first research question for the open ocean experiment is the degree of dependence of the regional scale features on the lateral boundary conditions versus the surface forcing (in this case the uniform SSTs in the regional model). This issue focuses on what Hans has called the "flushing rate" of the regional model. This experiment should be done without any nudging, of course. The ensemble idea suggested by Hans should be tested.
- The second question is can the dynamics and parameterized physics provide increased model skill? This is a lot to expect of the regional model, as it only has the parameterized (generally column model) physics with just the advection, Coriolis effect and pressure gradient forces on the regional model resolved scale available to process the incoming information from the lateral boundary and SST conditions. This question should be answered using real world data, not big brother/little brother experiments.
- When terrain or other surface features are present, we do see additional spatial structure, but this makes the assessment of the actual value added much more difficult than for the open ocean experiment.

5. The creation of different interior solutions with almost identical lateral boundary conditions is further evidence of the nonlinearity of the climate system. The use of ensembles with perturbed lateral B.C.s (and parameterizations, such as Barry has done and re initializations as Jeff and Liang have done) should be priority to propose in the paper that Liang is leading, and for a Workshop.
6. The methodology to evaluate value added will be central to this discussion. In my view, runs without any short of nudging represent an essential benchmark in such evaluations, as they illustrates the limitations in the regional models to create real world structure in the absence of external constraints except for the lateral boundary conditions. The assessment using Type 2 regional climate runs is needed in order to make statements in our paper on value-added from Type 3 and Type 4 regional climate model runs. Brier-score like concept is an excellent idea for such assessments with the Type 2 runs!
7. We do need to agree on how Type 1, 2, 3 and 4 dynamic downscaling are tested scientifically.

- For Type 1, this is done every day in weather forecasting, and we all agree on this approach.
- For Type 2, my recommended track is to use coarse reanalysis for the lateral boundary conditions and nudging and validate with finer scale reanalyses on an event basis. We should also continue to do the statistical evaluations that have been completed by several of us.
- For Type 3, we need to resolve, but I suggest we use seasonal forecasts where the SSTs are prescribed.
- For Type 4, suggestions are welcome, but he recommended that we use the IPCC model runs made for the last 20-30 years, and also for the next ten years.
- The fundamental requirement is that we should agree on a set of quantifiable tests of regional downscaling using real world data. I have concluded we need to do this on an event basis for Type 2 and Type 3 downscaling, with Type 4 have predictive skill that is necessarily the least of all three types.

8. In terms of value-added from RCMs (Type 2 and 3), he agrees that they can improve model accuracy compared with observations when the more detailed surface forcing forces mesoscale/regional weather patterns. We certainly see this, for example, when we used a reanalysis as lateral boundary conditions to force a RAMS over Florida. The reanalysis does not resolve the sea breeze. This is why your results (that you provide documentation for) improve on the larger scale model, since the surface forcing is important [incidentally, Tomislava Vukicevic has proposed using her adjoint work to quantify what fraction of the model results come from surface forcing and what fraction comes from lateral boundary conditions. However, this is not the appropriate test in his view. We need to perform the downscaling over ocean areas where there is minimal surface forcing on a spatial scale smaller than the larger scale model. If you can produce more realistic weather features on the regional domain than with the parent reanalysis (Type 2 run) and GCM (Type 3 run), this would be convincing. With respect to the use of the reanalysis, the power of this approach is that we have observed analysis data to compare with the RCM results. For instance, using the NARR to evaluate model performance when the lateral boundary conditions are from the NNR provides such a test. He does agree that Type 3 runs (with prescribed SSTs) such as performed with the COLA/ETA set are very effective. However, ETA can add value (accuracy) due to its finer spatial resolution of the surface forcing over land, in the example that you provi8ded. As he mentioned above, we need to evaluate this value-added over a region where surface forcing is not going to add information beyond that in the parent model. Finally, while he agrees with your thought experiment on the Type 2 experiments (i.e. that approximate lateral boundary conditions from the reanalysis, sampled from the real world, will create deviations in the interior of the "perfect" RCM), this deviation is what we assessed with our papers. In the Type 3 (and even more so in a Type 4 run), the lateral boundary conditions have no information at all on smaller scales than the parent model can resolve. Thus, they must provide poorer (i.e.

less accurate) lateral boundary conditions than when reanalyses are used to force the RCM. Even with a "perfect" RCM, how can the RCM correct for erroneous information that is being fed into the RCM? We concluded, therefore, that the Type 2 downscaling provides an upper limit on the accuracy of the RCM.

9. A fundamental question is how can the higher spatial resolution and more detailed parameterizations "correct" errors in the lateral boundary conditions if the internal regional model dynamics and physics significantly depend on the lateral boundary conditions? He proposes that we need to look at this issue over the ocean with no SST gradients. He agrees that when surface forcing is better resolved, that the regional model can still provide added skill, providing the errors in the lateral boundary conditions do not exert too large of an effect on the internal model results (thus directly affecting how the effect on the atmospheric dynamics and physics of the surface forcing is simulated). The skill added by the COLA/RCM appears to be due to the more spatially resolved surface forcing. We need to test this to see which of our views is correct.
10. IF the precursors of a jet stream over the Rockies and the low-level jet from the Gulf can be realistically simulated, the RCM could improve on this. However this is the crux of the issue. The coarse model must be able to skillfully insert the winds, temperature and moisture through the lateral boundary conditions as a necessary condition for the skillful RCM run with these synoptic features. Without this, how can the ETA RCM skillfully produce (and improve on) these synoptic features if the input of larger scale synoptic information is flawed? Clearly, it can only add value (skill) if the surface forcing is a dominate effect despite poor synoptic features. On the tests over water, our studies suggest that ALL of the RCM skill comes from more detailed terrain and landscape information, as you mostly agree with in your e-mail. This is why he suggests a test over open water as we can isolate the value-added (if any) with respect the ability of the RCM to improve on the skill of a coarse GCM (or Reanalysis). This would show whether or not the RCM can add skill at both the GCM (or Reanalysis) resolved scale and the finer scale. We need to run Type 2 and 3 experiments to test this.
11. The IPCC and US National Assessments appropriately should be communicated as process studies in the context that they are sensitivity studies. It is a very convoluted argument to state that a projection is not a prediction. The specification to periods of time in the future (e.g., 2050-2059) and the communication in this format is very misleading to the users of this information. This is a very important distinction which has been missed by impact scientists who study climate impacts using the output from these models and by policymakers.
12. In terms of the IPCC process, this would mean the starting point would be the threats to society and the environment from the spectrum of human and natural effects, and not by starting from the GCMs and downscaling. With the use of conditional forecasts (scenarios) the problem is that they may not capture all of the potential risks.

- Points rose by Barry:
  1. The additional problem we encounter, regardless, is that the regional model simulation is itself dependent not only on the large scale boundary conditions but also on how it "converts" those boundary conditions into meteorology (by choice of model parameterizations, etc).
  2. Global model might be able to reproduce characteristics of the general circulation on average, but can they realistically produce the higher frequency (and smaller scale) synoptic meteorology that drives regional model solutions and mesoscale model weather? If they cannot, it may not help us to use a general circulation at all to do dynamic downscaling, at least how they are currently used.
  3. Higher resolution simulation clearly adds spatial resolution of the meteorology.
  4. The meteorology at the higher resolution scale should be more accurate because the higher resolution domain resolves the underlying topography (including land elevation and surface type). This information is included in the 4 km simulation, but would not be included in a 45 km resolution simulation.
  5. "Good climatology consists of good meteorology:" The conclusion implies that any type of downscaling (type 2, 3, and 4) would fail unless the global model also simulates the meteorology, as well as climatology.
  6. The larger the RCM domain, the greater the likelihood it will be able to create realistic meteorological features (just imagine an RCM at high resolution extended over more than one continent, etc.)
  7. A GCM may send poor data into the boundaries, but if the RCM domain is large enough, it may still be able to develop more realistic synoptic meteorology in the interior of the domain -- since it is simulating higher resolution meteorology that can organize itself into larger scales.
  8. When making a seasonal simulation using climate model data, a larger RCM domain allows the RCM to spin up more realistic meteorology (see Feder's letter how underlying topography can help here).
  9. An RCM may never be able to faithfully reproduce reanalysis data, but over the long term it can produce more realistic meteorology than a GCM, producing a better regional product.

- Points rose by Christopher:
  - We need a large enough domain for RCM simulation to be physically useful, and the problems that are observed with loss of kinetic energy at large scales (e.g. Castro et al. 2005, Rockel et al. 2007) can be probably be traced back to the specification of the lateral boundary forcing. He disagrees with respect as to why Type 2 experiments are substantially different than Type 3 experiments. Only the type of lateral boundary forcing is being changed (from a "perfect" reanalysis to a GCM). It would be very straightforward to see if the same type of behavior (e.g. loss of kinetic energy at large scales) is observed in a Types 3 and 4 downscaling

modes. Simply force the RCM with data from a GCM, such as that of a NOAA CFS seasonal forecast (Type 3) or an IPCC projection simulation with an AOGCM (Type 4). Then do the similar analysis as He did in our 2005 paper and see if the behavior is equivalent.

- Points rose by Fedor:
    - He believes that the Type 2 experiments are not very useful in assessing the RCMs performance at large scales. To see why, consider a thought experiment in which we have a perfect RCM, performing as the real atmosphere. We are now driving this perfect RCM with the reanalysis LBCs which although sampled from the real atmosphere are sampled with an error; compared to the real atmosphere they are only approximate. While in a run long enough our perfect RCM will forget its initial condition, it will not forget the approximate LBCs that are constantly fed to it. To quote Ed Lorenz (acknowledgment: posting on Eugenia Kalnay office's door at UMD) chaos: when the initial condition determines the future but approximate initial condition does not approximately determine the future. Approximate LBCs will inflict the same behavior on our perfect RCM, the more so the bigger its domain is. Thus, our perfect RCM will not pass the test of emulating the reanalysis "truth", which does approximately describe the "future" of the real atmosphere.
    - It is in Type 3 and Type 4 experiments that tests free of the above discussed Type 2 difficulties can be done. An RCM driven by forecast GCM fields then has a chance to produce large scales with verifiable value added compared to those of the driver fields. To demonstrate that something IS possible already a single (believable!) experiment suffices, a task much easier than that that of demonstrating that something is not possible. Of course, some territory for argument remains, such as that value added is possible only due to a major weakness in some sense of the driver model. But even so, the demonstration of RCMs creating value added in large scales has been made, provided you accept that plots can serve that purpose.
    - Ruby Leung gave a talk at the NCAR regional climate modeling workshop two and a half years ago. She and her postdoc were running the same 1993-1988 case, using the NCAR WRF as an RCM, and had results that were less than satisfactory.
    - "More detailed surface forcing" include topography as surface forcing. This is where he believes most of the value added of the Eta RCM came from. Thus, he predicts it would not do all that well in Roger proposed tests over water.
    - "How can the RCM correct for erroneous information that is being fed into the RCM?" In principle he'd say it can, e.g., if it is handling the interaction of large scale flow with major topography (Rockies or Andes) significantly better than the driver GCM data describe it. But if so, this cannot be detected in Type 2 downscaling.

- With the same quality of the LBCs, the LBCs will indeed become “progressively less important” as the domain gets larger. “Important” may be a bit strong word here, certainly LBCs always are important and do matter; it is that the damage inflicted by the LBC errors becomes progressively somewhat smaller.

- Points rose by Hans:
  - The idea of correcting large-scale errors by the use of better resolution in a RCM has long, and unsuccessfully, tried by Bennert Machenhauer.
  - When talking about added value, one needs two well defined references Feser has used as references the DWD analyses (which should be replicated) and the NCEP reanalyses (compared to which the tested model must be better). Something like a Brier-score.

- Points rose by Dev:
  - Temperature accuracy is good but not critical unless it is for major changes in below freezing and frost occurrence probability. The users seem to be ok with large errors in temperatures away from freezing.
  - Both for temperature and rainfall, they were quite satisfied to only know a range of how much will it be below normal or above normal for the season. Further they desired this at monthly resolution. An error or up to 3F seemed acceptable! So the user requirements are in some sense very modest as they look to us for information beyond weather forecast as they good make operational decisions.

- Points rose by Yongkang:
  - How to evaluate the RCM results. We have to have some quantitative means to assess RCMs and define what "add value" means. He thinks in addition to RMS, correlation, Equitable Threat Score, the spectrum analysis, which was described in Castro et al (2005) and used in our paper, show very promising. As to the spatial structure, he is not that sure if there are any better methods other than spatial correlation. ETC and spectrum analysis also implicitly provide information in this aspect.
  - An article first to provide a better review on RCM issues. Filippo's review is too much one side story.
  - Physical parameterizations, such as convective schemes and land surface processes, always have big impacts. I think over Indian monsoon regions, these processes are crucial. The LBC can not replace basic dynamics and physics in the model.
  - The discussion about “good climatology consists of good meteorology” implies a serious RCM downscaling issue. He believes the statement is very good. But if we apply this statement to the model, we will conclude none of GCMs will be qualify for climate study because none of current GCM can simulate “meteorology” (as defined by Barry) well. His feeling is other way around more likely to be true, i.e. the model capable to simulate meteorology well should produce good climate.

- The RCM climate downscaling intends to have both meteorology and climate be well simulated. If RCM is not able to add value, then type 2, 3, and 4 downscaling will deem to fail, and statistic downscaling is more meaningful. From some current RCM studies, he thinks one issue needs to be considered, i.e., whether the added value is due to RCM's high resolution and/or better topography. If this is the case, then RCM will have limited future since more and more high resolution GCMs will be in–use.
- Reanalysis LBC versus GCM LBC. RCMs suppose to provide small scale features which are NOT exist in LBCs. Otherwise, there is no need to have RCM's downscaling. So the question is under what kind of LBCs, a perfect RCM can properly conduct downscaling. It seems one consensus is that Reanalysis LBC may be able to produce proper downscaling but not GCM (with less consensus). Since both global reanalysis and GCM have NO small scale information, the questions would be how good large scale features should be for a proper perfect RCM to conducting proper downscaling. If we give the quality of global reanalysis LBC a scale 90-100 (since there are different reanalyses with different quality), the question would be for a GCM, at what scale, (say 80, 50, etc?), their LBC are possible to help RCM to add new (small scale) value, or not at all. Here the scale is just an example; more proper criterion could be "proper jet stream", etc.
- The domain size. When the RCM domain becomes larger and larger, it should be able to generate more and more independent information from LBCs. The largest RCM is a GCM. The question is whether a very high resolution GCM is able to properly produce both large scale and small scale features. Very high resolution GCMs already exist and have generated quite a few results. I have not done a thorough investigation on them, but from some results I saw, it seems to me RCMs could still survive for a while. Now the question is how large the RCM domain should be. Our recent study (Xue et al., 2007, J. Climate) has indicated that when the RCM domain gets larger, the internal noise is so large which produced significant regional climate drafting. Therefore, although we wish to have a large domain for RCM to generate correct local information, but it should not be too large because of the climate drifting /noise, etc. A proper domain seems important but the real size may very much depend on individual model.