CCMC MEETING NOTES: SESSION ON HARDWARE, SOFTWARE, & DATA ASSIMILATION (Wednesday morning session)

Executive summary:

- **Hardware issues**
  - There are a lot of hardware resources available for use by CCMC and code developers
  - There may be issues of access to the DREN network to get to these resources
- **Software issues**
  - There is a lot of support software (developed by PACI) available. CCMC is working with MHPCC staff to arrive at a reasonable solution

- **Data assimilation issues**
  - Data assimilation is very difficult, and is currently in a basic research mode, and not mature enough to be included as part of CCMC activities
  - Data validation is a reasonable activity for CCMC
  - Data validation will be a necessary step before data assimilation is eventually included as a CCMC activity

Detailed notes on session talks:

Hardware Assets for Space Weather Modeling

Ariel Acebal provided a summary of computing hardware located at the AFWA facility in Omaha, Nebraska. He mentioned the GAS/SWAFS facility, consisting of 12-nodes with a 45 Gflop capability, connected by DREN. He also mentioned the current CCMC cluster, with 70 nodes and a 44 Gflop capability, that will soon be replaced by a system at GSFC.

Tim Fahey then described the MHPCC, a distributed center, which is used primarily for research and development, with a small part of the center dedicated to support for the Air Force Observatory on Haleakala. Currently, resources available to CCMC and space weather code developers include “Tempest”, with with 53 nodes, with 16 CPU/node, and “Huinalu”, an IBM linux supercluster with 520 933 MHz processors. There are 2 CPU per node, connected with high speed switches. These are unclassified machines accessible over DREN. DREN bandwidth is currently T3, but going to OC3 fairly soon.

Terry Onsager of the NOAA/SEC then described the computing hardware available at NOAA/SEC. SEC itself does not run huge models. They do have a Beowulf cluster with 16 nodes, plus a broad spectrum of other hardware. The Beowulf cluster is used for research purposes and is not used operationally. SEC has access to NSEP IBM SP-2s, as well as the JET 256 node Beowulf system. Terry was not sure whether these latter facilities are available to outside users.

Steven Quigley of the Air Force Rapid Prototyping Center (RPC), whose mission it is to obtain models for operational use for delivery to ASWA, described the RPC's facilities. He mentioned in particular a SUN Sunfire 4800 with 8 processors, intended primarily for internal usage, but could be made available to CCMC.

Jimmy Raeder of UCLA then described his experiences with Beowulf clusters. He first recommended that anyone interested in building a cluster first look at the website http://www.beowulf.org and the links therein. Beowulf clusters consist of an array of Linux machines connected over a local area network that can be used to solve large scale numerical problems (the movie "Toy Story" was rendered on a 140 node cluster). The "poor man's solution" is to use cheap 100 Mb s^{-1} switches between the nodes, which gives a bandwidth that is a factor of 10-20 less than an IBM SP. However, this does not always affect code performance seriously. A step up from the "poor man's solution" is the use of 1Gb s^{-1} switches, which makes the connections faster, but the latency is not significantly improved (80 μs versus 100 μs). Raeder then showed a plot of the performance per processor for Beowulfs versus several other types of computing hardware as a function of grid points per node along one direction in a 3d simulation. Beowulf clusters show a ``sweet spot'' at about 15 gridpoints, with the AMD Athlon processor having the top performance. When the number of gridpoints (along one direction) exceeds 60, Beowulf performance suffers compared to the IBM SP, and becomes comparable to the SGI Origin. Raeder also compared the performance of various compilers, with the top performer being the GNU g77 Fortran compiler. MPI is used to share information between nodes, with 2 common implementations being MPICH and LAM, both freely available. Raeder recommended that anyone contemplating a Beowulf not skimp on the motherboard, and to go to websites frequented by "over-clockers" to find the most robust motherboards.

Software assets for CCMC

Chuck Goodrich (U. Md) gave a summary of the job management software developed by the PACI centers in order to support the creation of the national (computational) grid and the developments in the supercomputing program at the NSF which led to their development. He introduced one promising package, GLOBUS. Michael Hesse announced afterward that the CCMC was exploring GLOBUS now. Ayris Falasca of the CCMC met subsequently with the expert in GLOBUS at MHPCC.

Data Assimilation in Space Weather

George Fisher gave a brief presentation on the issues of driving Solar MHD simulations with observed data from vector magnetographs, with his main point being that MHD models require a self-consistent specification of the magnetic field and velocity, whereas the data will most likely consist of just arrays of magnetic field vectors at the photosphere. He described how it might be possible to derive the velocity field from the observations provided one can assume that ideal MHD obtains at the photosphere. Following the talk there was discussion about the possible utility of Kalman filters and other formal assimilation techniques, as well as the difficulties associated with noise in the transverse field measurements.

Jeff Kuhn of UH then described how one can measure magnetic fields in the corona directly, using magnetically sensitive forbidden lines formed in the infra-red part of the spectrum. To convert these measurements into magnetic field maps of the corona, one must de-convolve the magnetic field from line of sight integrals that give the measured polarization signal above the solar limb. He described some ideas for how this might be done in the future.

Zoran Mikic of SAIC outlined the use of filtered, large scale magnetic field maps from e.g. Kitt Peak magnetograms, as the boundary conditions for MHD simulations to determine the large scale structure of the solar corona that one might see during a solar eclipse. Their code has been used to make predictions of what will be seen during solar eclipses for several years.

In the Heliosphere, one can make measurements relevant to large scale structure through the use of Thomsen scattering of white-light from the Sun, or through the use of radio scintillation of far-away radio sources. Bernie Jackson of UCSD described and illustrated with examples how both kinds of measurements can be converted into 3-D time dependent maps of the Heliosphere through the use of tomographic reconstruction. His group at UCSD is collaborating with Detman of SEC to combine this work with MHD simulation to compute magnetic field evolution in the solar wind.

John Lyon of Dartmouth discussed data assimilation as a means to produce more accurate magnetospheric models. The most promising areas for assimilation appear to be in the inner magnetosphere and situations, like the radiation belts, where the dynamics of the global system don't depend on the modeled quantity. There are also opportunities to assimilate interior magnetospheric observations to modify the solar wind driver within observational limits.

Bob Schunk assessed the state of data assimilation into ionospheric models. He pointed out the large array of observations amenable to assimilation and the challenge in incorporating these data into the models. The Utah State effort is evaluating in parallel a number of sophisticated techniques for this assimilation.