PANEL DISCUSSION

Panel 2: Evaluation of Models

Moderator: Dr. David Rogers, *Director, Office of Weather and Air Quality, Office of Oceanic and Atmospheric Research*

Rapporteurs: Mr. Thomas Fraim, *Senior Meteorologist, Office of the Federal Coordinator for Meteorology*

Mr. James McNitt, *Office of the Federal Coordinator for Meteorology (Science and Technology Corporation)*

Synopsis

Dr. Rogers opened the session with comments regarding the importance of evaluating models in the real environment with user criteria. He noted the need to establish a comprehensive urban observing system for model evaluation that the federal agencies would fund. He proposed a test-bed in the Washington, DC area.

Dr. Rogers then introduced three speakers who discussed model evaluation processes in their respective federal agencies. An opportunity for questions and discussion followed these presentations.

A question was asked about the availability of urban databases. The DOD panel member stated that for purposes of DOD's interim model accreditation, there are three data sets deemed sufficient for urban evaluation. These are the Salt Lake City data set, the Los Angeles data set used for MIDAS AT, and the UK data set used for the Urban Dispersion Model (UDM) evaluation. It was noted however, that additional urban data will be required in order to fully accredit the DOD's new Joint Effects Model (JEM). The planned Oklahoma City experiment in July 2003 will provide an opportunity to gather more data in an urban environment. Following a comment that the Los Angeles data set consists of SF_6 releases in downtown Los Angeles over a several day period, the EPA panelist emphasized the point that in the future, longer time period data sets will be needed in order to deal with evaluation statistically. In that regard, he stated that Project Prairie Grass was a well designed experiment the data from which has withstood the test of time.

In keeping with the idea of longer period data sets, there was some discussion on the notion of a long-term study in the Washington, DC area. This would allow for the development of probability distributions rather than just discrete concentrations in space and time. Another point of view was raised from the audience with respect to probability distributions. The point was stressed that the user generally wants a definitive answer and isn't real interested in probabilities. It was pointed out that perhaps there is a communication problem and it might be better to build the uncertainties into the definitive forecast rather than passing them on to the user and leaving the interpretation up to him. However, in order to do this, the user must communicate his thresholds to the forecaster so the final answer incorporates all user

sensitivities. It was noted by the DOD panel member that the DOD's Chemical Stockpile Emergency Preparedness Program (CSEPP) has worked with states and counties to establish thresholds and that within minutes of an event using D2PC, information is available for the decision maker.

A question arose regarding what criteria might be used for deciding whether models are performing satisfactorily. There was some discussion as to whether a Figure of Merit (FOM) or some other measure such as median or mode should be used. The DOD panelist stressed the need to balance "falsely warned" with "falsely unwarned" (i.e., false alarms with missed events) and that the user should decide his tolerance for one or the other. It was also noted that the DOD accepts a certain degree of risk that might not be acceptable with civilian populations. The EPA panelist noted that in urban situations the models are showing a great deal of sensitivity to multi-scale issues (e.g., building morphologies). In this regard, it was emphasized that models should be evaluated for their intended purpose and that the evaluations should be organized with respect to the appropriate scale(s).

The panel moderator brought up the notion of a continuous evaluation of the models running in real-time. For example, NOAA is running HYSPLIT continuously for a number of locations, so ample model data would be readily available for a statistical evaluation. The DOD panelist responded that this might be possible for a limited number of scenarios, but for the DOD, the number of possible combinations is too numerous. Rather than this approach, the DOD would like to implement nowcasting so that real-time weather is available soon after an event happens. The DOD is also looking at denser weather data networks to benefit various applications including transport and diffusion. The EPA panelist mentioned the need for a continuous source of some kind to make such evaluations viable. He also mentioned that the EPA and NOAA are embarking on a program to continuously forecast ozone. This effort might benefit the continuous running of models for other applications including transport and diffusion. A member of the audience highlighted the need to develop the capability to examine the flow field using remote sensing, especially satellite and radar. This would provide supporting weather data in areas where mesoscale observing networks have not been deployed.

The moderator wrapped up the discussion by emphasizing a point that had come up several times during the session: the user must be engaged in any development of new tools and capabilities.