Do Training Datasets Make a Difference?

Harold M. Woolf
University of Wisconsin-Madison/CIMSS

Members of the remote sensing community, and especially those working in the area of atmospheric-profile retrieval science, often expend considerable energy and resources in the construction of training datasets for two of the primary components of that work: (1) fast transmittance models, and (2) regression first-guess schemes. Recently the author has begun to wonder just how significant those efforts are in terms of the end result -- the quality of the retrieved profiles of temperature and humidity.

A study has been undertaken, utilizing the IAPP and the flyover, or direct-readout, NOAA-15 and -16 ATOVS datasets received at Madison, Wisconsin, to assess the influence of different training datasets on retrieval accuracy. In the area of fast transmittance model generation, the CIMSS-32 (31 profiles plus the Standard Atmosphere) and UMBC-49 (48 profiles plus the Standard Atmosphere) training datasets have been employed to construct coefficients for the PLOD/PFAAST algorithm. For regression first-guess development, the NOAA-88b and SEEBORv3 datasets have been used.

Retrievals have been produced from all available flyover passes since 10 February 2005, using the following "combinations of ingredients":

A. the existing "operational" version, consisting of CIMSS-32 fast transmittance and NOAA88b first guess;
B. UMBC-49 fast transmittance and NOAA88b first guess;
C. UMBC-49 fast transmittance and SEEBORv3 first guess.

For all three versions, the retrievals are also run with NWP first guess.

Assessments of retrieved temperature and humidity quality, based on radiosonde matchups, are presented for the six cases, and conclusions are drawn as to the significance of the choice of training datasets for the two components.

INTERNATIONAL ATOVS WORKING GROUP

Proceedings of the Fourteenth International TOVS Study Conference

Beijing, China
25-31 May 2005