Do Training Datasets Make a Difference?

Harold M. Woolf

Cooperative Institute for Meteorological Satellite Studies (CIMSS)

University of Wisconsin-Madison

The Issue

Members of the remote sensing community, especially those working in atmospheric-profile retrieval science, often expend considerable energy and resources in the construction of training datasets for two of its primary components:

(1) fast transmittance models;
(2) regression retrieval.

While the latter often provides a first guess for some physical retrieval technique, it can also serve as the "answer."

The Experiment

An experiment was designed to employ the IAPP retrieval package and the "flyover" (direct-readout) ATOVS passes received at Madison, Wisconsin, to assess the influence of different training datasets on retrieval accuracy. At the same time, the effect of advances in the underlying line-by-line calculations on the performance of fast transmittance models, along with changes in the training datasets, could be evaluated.

In the area of radiative transfer, the CIMSS-32 (31 profiles plus the Standard Atmosphere) and UMBC-49 (48 profiles plus Standard Atmosphere) training datasets have been employed to construct coefficients for the fast models. For regression retrieval/first-guess development, the NOAA-88b and SEEBORv3 datasets have been used.

Retrievals have been produced from all available NOAA-16 flyover passes from 10 February 2005 onward, using the following "combinations of ingredients":

A. existing "operational" version (IAPPv2.01): CIMSS-32 fast transmittance, NOAA88b first guess (Old/Old);
B. UMBC-49 fast transmittance and NOAA88b first guess (New/Old);
C. UMBC-49 fast transmittance and SEEBORv3 first guess (New/New).

For all three fast-model/regression situations, the retrievals were also produced with the aid of an NWP first guess, obtained from NCEP's GFS (formerly AVN) model.

The fast transmittance algorithm employed in all cases is PLOD or PFAAST (Hannon et al., 1996). The coefficients used in (A) were determined from line-by-line calculations with LBLRTMv6.01 (Clough and Iacono, 1995) and HITRAN1996/JPL_EXT (Rothman et al., 1992) for HIRS, and with the procedures of Rosenkrantz (1975) and Barrett and Chung (1962) for AMSU. For those used in (B) and (C), infrared computations were performed with LBLRTMv8.4 and HITRAN2000/AER_UPDAT_01.1, while those for the microwave were done with the Liebe (1993) model. In each case, the regression retrieval coefficients were derived from simulated brightness temperatures obtained via forward calculations with the corresponding fast transmittance model.

Results and Conclusions

As of 15 May 2005, after the experiment had been running for 95 days, the total number of retrieval-radiosonde matches was 57483. After extensive quality control to screen out unrealistic radiosonde data and obviously bad retrievals, the resulting sample of 25326 matches – which decreased to 23525 at 850 hPa, 18649 at 950 hPa, and 7693 at 1000 hPa due to the elimination of soundings over elevated terrain – was subjected to the usual statistical analysis to determine mean and root-mean-square differences between retrieval and radiosonde values of temperature and water-vapor mixing ratio.

The results of that analysis, presented below in four composite figures, suggest that the incorporation of more up-to-date radiative transfer and regression training has not made a major impact on the accuracy of IAPP retrievals. One exception is the bias, or mean retrieval-radiosonde difference, in water-vapor for the regression first guess (upper right figure), in which the new regression training combined with the new radiative transfer (Mean New/New) shows marked improvement over the other combinations. It is also noteworthy that when the NWP first guess is employed, the results for all three situations are virtually the same. This suggests that the current retrieval algorithm is perhaps not allowing optimal changes to the first guess. This will be the subject of further investigation.

References

Barrett, A.H. and V.K. Chung, 1962: "High-altitude water vapor abundance from ground-based microwave observations." Journal of Geophysical Research, 67, 4259-4266

Clough, S.A. and M.J.Iacono, 1995: "Line-by-line calculations of atmospheric fluxes and cooling rates. 2: Applications to carbon dioxide, ozone, methane, nitrous oxide and the halocarbons." Journal of Geophysical Research, 100, 16519-16535

Hannon, S., L.L.Strow, and W.W.McMillan, 1996: "Atmospheric infrared fast transmittance models: A comparison of two approaches," in Proceedings of SPIE Conference 2830, Optical Spectroscopic Techniques and Instrumentation for Atmospheric and Space Research II, 12pp.

Liebe, H.J., G.A. Hufford, and M.G. Cotton, 1993: "Propagation modeling of moist air and suspended water/ice particles at frequencies below 1000 GHz".In AGARD Conference Proceedings 542, Atmospheric propagation effects through natural and man-made obscurants for visible through MM-wave radiation, 3.1-3.10

Rothman, L.S., R.R.Gamache, R.H.Tipping, C.P.Rinsland, M.A.H.Smith, D.C.Benner, V.Molathy Devi, J-M.Flaud, C.Camy-Peyret, A.Perrin, A.Goldman, S.Massie, L.R.Brown, and R.A.Toth, 1992: "The HITRAN Molecular Database: Editions of 1991 and 1992," Journal of Quantitative Spectroscopy and Radiative Transfer 48, 469-507

Rosenkrantz, P.W., 1975: "Shape of the 5mm oxygen band in the atmosphere." IEEE Transactions of Antennas and Propagation, AP-23, 498-506