The Application of Principal Component Analysis (PCA) to AIRS Data Compression

Lihang Zhou[1], Mitchell D. Goldberg[2], Walter W. Wolf[1], and Chris Barnet[2]
[1]QSS Group, Inc
[2]NOAA/NESDIS/ORA

Data compression is one of the key issues for high spectral resolution infrared sounders because their data volumes are too large to distribute in full to the users and the data assimilation centers. The Atmospheric Infrared Sounder (AIRS) (Aumann et al. 2003), launched on May 4, 2002 on the AQUA-EOS satellite, is the first of a new generation of high spectral resolution infrared sounder having 2378 channels measuring outgoing radiance between 650 cm-1 and 2675 cm-^{1}. NOAA/NESDIS is processing and distributing AIRS data and products in near real-time to operational NWP centers. This offers us a great opportunity to use real AIRS observations as a test-bed for data compression study of the hyper spectral sounding instruments.

The desired features of hyper spectral data compression include high compression ratio, fast processing time, and the preservation of relevant information. Principal Components Analysis (PCA) provides an effective way to reach these goals. Since the information from the 2000 plus AIRS channels are not independent, PCA can be used to reduce the dimension while retaining the significant information content of the data. The AIRS spectrum can be represented by a much smaller amount of PCA scores. Individual channels can be reconstructed with minimal signal loss. Reconstruction errors can be coded using Huffman coding. Instead of the individual channel radiances, principal component (eigenvector) coefficients can be provided to the users, thus reducing the size of the data volume. The coded reconstruction errors and the corresponding statistical metadata can be provided along with the data if lossless compression is desired.

Our preliminary study shows that compression factors of up to 50 can be obtained with this approach, without losing any accuracy of the data. In NOAA NESDIS, a data compression system based on this approach is being developed, which will allow us to archive and distribute the compressed AIRS level 1B data and the corresponding metadata in near real time. In this poster we present our studies of the application of this approach to other AIRS data compression techniques. The generation and application of the eigenvectors, the process of Huffman coding the reconstruction errors and the creation of the corresponding statistical metadata will be described. The knowledge that we are gaining from the AIRS data will be useful for high spectral infrared radiance data compression for future satellite observations.

INTERNATIONAL
ATOVS
WORKING GROUP

Proceedings of the Fourteenth International TOVS Study Conference

Beijing, China
25-31 May 2005