

Intercomparison of Climate Data Sets as a Measure of Observational Uncertainty

by

Curt Covey
Krishna M. AchutaRao
Michael Fiorino
Peter J. Gleckler
Karl E. Taylor
Michael F. Wehner

Program for Climate Model Diagnosis and Intercomparison
Lawrence Livermore National Laboratory

April 2002

DISCLAIMER

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

This report has been reproduced
directly from the best available copy.

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information
P.O. Box 62, Oak Ridge, TN 37831
Prices available from (615) 576-8401, FTS 626-8401

Available to the public from the
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Rd.,
Springfield, VA 22161

ABSTRACT

Uncertainties in climate observations are revealed when alternate observationally based data sets are compared. General circulation model-based "reanalyses" of meteorological observations will yield different results from different models, even if identical sets of raw unanalyzed data form their starting points. We have examined 25 longitude-latitude fields (including selected levels for three-dimensional quantities) encompassing atmospheric climate variables for which the PCMDI observational data base contains two or more high-quality sources. For the most part we compare ECMWF with NCEP reanalysis. In some cases, we compare in situ and / or satellite-derived data with reanalysis.

To obtain an overview of the differences for all 25 fields, we use a graphical technique developed for climate model diagnosis: a "portrait diagram" displaying root-mean-square differences between the alternate data sources. With a few exceptions (arising from the requirement that RMS differences be normalized to accommodate different units of variables) the portrait diagrams indicate areas of agreement and disagreement that can be confirmed by examining traditional graphics such as zonal mean plots. In accord with conventional wisdom, the greatest agreement between alternate data sets - hence the smallest implied observational uncertainty - occurs for upper tropospheric zonal wind. We also find fairly good agreement between reanalysis and more direct measures of precipitation, suggesting that modern observational systems are resolving some long-standing problems with its measurement.

Introduction

The improvement of climate models is limited not only by computational constraints and uncertainties in climate processes, but also by uncertainties in the observed climate record. Model simulations of the present and the recent past are routinely compared with observations in an effort to gauge model performance and to obtain some indication of the amount of confidence that can be placed in simulations of the future (e.g., McAvaney et al. 2001). Unfortunately, climate observations rarely come with comprehensive uncertainty estimates including systematic as well as random effects. Since any model simulation will deviate from observations by a finite amount, evaluators of models face the choice of either declaring agreement with observations "satisfactory" in a qualitative (and often poorly specified) sense, or branding the models as "erroneous" in one or more respects. This dilemma is especially acute in studies that attempt a broad overview of model performance.

Another complicating factor is that climate observations used in model validation generally come from an analysis of observational data to a space-time grid. The analysis process may be simple interpolation or as complex as four-dimensional data assimilation with a high resolution general circulation model ("reanalysis"). Regardless of the process, the final product is a gridded representation of a geophysical quantity based on irregularly spaced observations of the quantity itself (or of related quantities, e.g., satellite-obtained radiances). From this point forward we will refer to analyses of observations as observations while recognizing that the analysis process is a source of error that is difficult to quantify.

There are no simple solutions to these problems. Progress in deriving bounds for some components of meteorological and climatological observations has been made, e.g., by exploiting physical constraints such as conservation laws (Trenberth et al. 2001, 2002). Nevertheless, the variety and complexity of the data guarantee that rigorous and believable error bars will be missing from many of the numbers. In such cases the best alternative may be to compare different observations of the same quantity. Of course such a procedure will underestimate observational errors if the competing data sets are not significantly independent. It can also overestimate observational errors if one of the data sets is clearly inferior to the others. Often it will not be possible to ascertain which of these two effects is more important, and thus the intercomparison of climate data sets can serve as only a crude measure of observational uncertainty. In many cases it may nevertheless be the only available comprehensive measure.

The growing number of climate observations -- most notably from several model-based reanalyses of meteorological data, but also including new compilations from satellites and other sources -- allow climate data set intercomparison to be undertaken in fairly systematic fashion. In this report we do so using the observational data base of the Program for Climate Model Diagnosis and Intercomparison. It is our intention to place this data base in the public domain in user-friendly form; see Fiorino (2002) and <http://www-pcmdi.llnl.gov/obs/pods/mo/obs.mo.var.htm> We focus on root-mean-square (RMS) differences between competing data sets. This provides a simple quantitative measure of overall differences, although it undoubtedly misses some important patterns apparent to the human eye in traditional presentations such as latitude-longitude maps.

Data Sources

Table 1 shows the observed variables examined in this report and two data sources for each variable (including references for the data sources). We use all variables in the PCMDI observational data base for which two or more sources exist. These include most of the atmospheric variables of interest in evaluating climate model performance. We require that both data sources provide global coverage of at least three years duration, with monthly time resolution tied to specific calendar years (rather than a climatological seasonal cycle). This requirement enables us to examine interannual variations in addition to average climatology, but it excludes data sets that would otherwise be desirable such as COADS for surface fluxes.

For most variables in Table 1, the two data sources are reanalyses of meteorological data from the U.S. National Center for Environmental Prediction (NCEP) and the European Centre for Medium-range Weather Forecasting (ECMWF). The NCEP and ECMWF reanalyses use nearly the same raw observations but employ different numerical prediction models to interpolate the data to a regular grid and to make different variables consistent with theoretical expectations. Theoretical consistency imposes strong constraints on variables such as clouds, which are subject to rapid adjustment by the models. For such variables the reanalysis products are almost completely determined by the models, and the NCEP and ECMWF reanalyses can be considered independent. At the other end of the spectrum lie variables such as upper-air temperature, geopotential height and nondivergent wind, which are strongly influenced by the raw observations. Even in the latter cases, data must be interpolated to a regular grid in time and space, and techniques for doing so will vary between the models.

For a few variables -- cloudiness, precipitation, top-of-atmosphere energy fluxes, and surface air temperature -- we use data obtained more directly from in situ and / or satellite observations. We compare these "direct" observations with one of the two reanalyses or (in the case of outgoing longwave radiation) with alternative satellite data. It must be stressed that the "direct" observations, like reanalysis data products, arise from an analysis process that begins with raw data irregularly spaced in space and time, and often only indirectly related to the quantity of interest (e.g., satellite-detected radiances in narrow directional and frequency bands). From this point forward we refer to analyses of observations as observations while recognizing that the analysis process is a source of error that is difficult to quantify.

Subjective judgement inevitably influences our choices of data sets. For precipitation, we compare the CMAP data set with reanalysis rather than with the alternative from the Global Precipitation Climatology Project (GPCP; see Gruber et al. 2000) because CMAP and GPCP seem to be essentially the same data over land. For outgoing longwave radiation, on the other hand, we use two satellite-derived datasets -- ERBE and CPC -- that appear to us to be sufficiently independent since they come from broad-band and narrow-band sensors respectively. We also designate one of the two data sets for each variable as "primary" and the other as "secondary". The difference is not significant for the present study, but it will become important in future work, in which we will prefer the primary source when comparing model simulations with observations. We designate

ISCCP and ERBE satellite data as primary for cloudiness and top-of-atmosphere energy fluxes respectively. For surface air temperature, our primary source is IPCC / Jones, a merged land and ocean data set used in the most recent assessment report of the Intergovernmental Panel on Climate Change. For variables in which we compare ECMWF and NCEP reanalyses, we designate NCEP as primary since we believe it will be more familiar and accessible to most of our readers.

Presentation Format

Figure 1 shows space-time RMS differences between primary and secondary observations for each of the variables in Table 1. This figure is a color-coded table of values, or "portrait diagram", a type of diagnostic developed for model intercomparison (e.g., Fiorino 1995, McAvaney et al. 2001). For each field (column), different rows give RMS differences for different seasons and latitude bands, as indicated by labels on the figure. There are 12 single-level fields (surface air temperature, sea level pressure, etc.) and 7 three-dimensional fields. For 6 of the 7 three-dimensional fields, we show the RMS difference for 2 selected pressure levels, typically 200 and 850 hPa. Water vapor appears at a different set of levels, because relative humidity data is very inaccurate for $P < 500$ hPa and specific humidity data is absent for $P < 300$ hPa. The remaining three-dimensional field, geopotential height, is analyzed only at 500 hPa. There are 25 columns and 275 table entries in all.

Detailed background information for portrait diagrams is given in the Appendix. In brief, since the different variables have different units, comparing them with each other requires that their RMS differences be normalized to some uniform scale. We choose to normalize the differences for each field by the standard deviation of space-time variations of the primary observed field. We then color-code the normalized RMS differences: values increasingly greater than 1 are shown by increasingly deep shades of red, and values less than 1 by shades of blue. We take blue colors to suggest good agreement between primary and secondary data sources and red colors to suggest poor agreement. These choices are somewhat arbitrary, but they provide a convenient preliminary view of a large amount of data.

Figures 2-4 are additional portrait diagrams that subdivide the total RMS difference shown in Figure 1 into three parts. These are, respectively, the area and time mean (or "bias") component, the climatological seasonal cycle space-time pattern component, and the interannual space-time pattern component. The bias component has the same normalization as the total RMS difference shown in Figure 1, i.e., mean differences are divided by the total standard deviations of the reference fields. The climatological and interannual components, however, are normalized by the climatological and interannual variations of the reference fields respectively. As explained in the Appendix, the interannual components displayed in Figure 4 are such small fractions of the total RMS differences that normalization by total (or by climatological) reference field variations would make them invisible.

The remaining figures present our results in more conventional formats. Figures 5-19 are one-dimensional plots of zonal means of selected fields from Figures 1-4. (The order of Figures 5-19 is chosen to match the order of the fields, from left to right, in Figures 1-4.) Each of Figures 5-19 shows climatological June - July - August (JJA) and December -

January - February (DJF) means for both primary and secondary observations, together with the differences between the two classes of observations. The final figures are two-dimensional graphics in latitude-longitude or latitude-pressure space.

Our discussion below is guided by the portrait diagrams (Figures 1-4), making reference to each of the remaining figures as the occasion arises.

Results and Discussion

Figure 1 indicates that the best agreement between primary and secondary data sources occurs for zonal wind (ua). In this case the normalized RMS differences between the data sources are < 0.6 for both pressure levels, and for all seasons and latitude bands, shown in the figure. The corresponding zonal mean plots (Figures 17-18) show differences between the data sources that are very small compared with Equator-to-Pole variations. Nearly as good levels of agreement, according to Figure 1, occur for surface air temperature (tas), air temperature (ta) at 850 hPa, sea level pressure (psl), outgoing longwave radiation at the top of the atmosphere ($rlut$), geopotential height (zg) at 500 hPa, and meridional wind (va) at 200 hPa. In these cases the normalized RMS differences between the data sources are always < 0.8 and usually < 0.6 . The zonal mean plots for tas (Figure 5), ta at 850 hPa (Figure 16), psl (Figure 6), and zg at 500 hPa (Figure 14) support the inferences from Figure 1 by showing very small differences between the primary and secondary data sources. (Exceptions occur over Antarctica, where all data sources and in some cases definitions of the variables are problematic.) For ua , ta , psl and zg at 500 hPa, we are comparing the ECMWF and NCEP reanalyses, which start from basically the same set of observations; for these variables the models may do little more than interpolate to a regular grid. The fact that tas also shows good agreement between primary and secondary data sources is more surprising. In this case we are comparing NCEP reanalysis with in situ thermometer data. The model has more "scope for the imagination" in processing near-surface temperatures, yet agreement with in situ data is very good according to Figure 1.

Differences between primary and secondary data sources are more apparent in zonal mean plots for $rlut$ (Figure 9) and va at 200 hPa (Figure 19), despite the implication of Figure 1 that agreement is good. Similar statements can be made for precipitation (pr) and vertical motion (wap) at 200 hPa. Figure 1 implies that agreement between the data sources is as good or better for pr than it is for wap at 200 hPa, but this conclusion is not supported by examination of the corresponding zonal mean plots (Figure 7 and Figure 11). The case of precipitation is particularly surprising. Figure 1 shows that normalized RMS differences between the data sources for pr (the Climate Prediction Center merged analysis [CMAP] and NCEP reanalysis) are < 0.8 for all seasons and latitude bands examined. By this measure precipitation is among the better observed climate variables, in contrast with its notorious reputation. In this case the portrait diagram may be underrating differences between test and reference data. Precipitation exhibits large natural climate variations in space and time, and use of its natural variability for normalization reduces our measure of data source differences. On the other hand, a recent comparison employing more traditional metrics (Kistler et al. 2001, Figs. 16 and 18) shows several modern precipitation data sets in fairly good agreement, at least for their spatial mean and climatological components.

Turning to fields for which Figure 1 indicates poor agreement between primary and secondary data sources, we note that upper tropospheric humidity (*hus* at 300 hPa) stands out in this regard. Normalized RMS differences between the data sources (ECMWF and NCEP reanalyses) are of order 1 or greater for nearly all seasons and latitude bands examined. Figure 2 and Figure 3 show that the space-time mean (or "bias") component dominates the total RMS differences, especially in the tropics. For other fields, in contrast, the dominant source of RMS differences is typically space-time variations associated with the climatological seasonal cycle ("climatological component"). (Figure 12) shows that NCEP reanalysis gives much more upper tropospheric humidity than ECMWF reanalysis, confirming the dominance of the "bias" component.

Total cloudiness (*clt*) also exhibits fairly large RMS differences between its primary (ISCCP) and secondary (NCEP reanalysis) datasets. In this case the "bias" component is less important than space-time variations in contributing to total RMS differences. Consistent with this implication of the portrait diagrams, Figure 8 shows a poor correlation between the the two data sets. Here we use the latest version of NCEP reanalysis, since the original version contained errors that underestimated cloudiness; despite this correction, NCEP reanalysis gives generally less cloudiness than ISCCP and the "bias" component of the RMS difference is not negligible. It is also noteworthy that if we compare ECMWF rather than NCEP reanalysis with ISCCP for cloudiness, the RMS differences are noticeably reduced (*result not shown*). This may be due to ECMWF's more extensive incorporation of satellite-based radiances.

A few fields appear from Figure 1 to have poor agreement between primary and secondary data sources at just a few of the space-time subdivisions (rows), and relatively good agreement elsewhere. Most spectacular in this regard is air temperature (*ta*) at 200 hPa. This field has normalized total RMS differences > 1.6 in the tropics and < 0.8 elsewhere. Figures 2-3 indicate that most of these differences are associated with the bias. The corresponding zonal mean plot (Figure 15) shows very little mean difference compared with Equator-to-Pole variations (*top graph*), although the differences are systematic with ECMWF reanalysis ~ 1 K cooler than NCEP reanalysis at low- and mid-latitudes (*bottom graph*). Reflected solar radiation at the top of the atmosphere (*rsut*) is a similar though less spectacular case: normalized total RMS differences are > 1 in the tropics, and in the March - April - May season globally, but < 0.8 elsewhere. The corresponding zonal mean plot (Figure 10) shows that equatorward of $\pm 30^\circ$ latitude, NCEP reanalysis gives $10\text{-}20 \text{ W m}^{-2}$ larger values than ERBE. Finally, for lower tropospheric humidity (*hur* at 850 hPa), both Figure 1 and Figure 13 give some indication of greater disagreement between the data sources in the June - July - August season than in the December - January - February season.

The case of air temperature deserves closer examination. As noted above, temperature above the surface boundary layer is among the variables least influenced by model-based processing (the so-called type A variables in Kalnay et al. 1996). It may seem surprising that the portrait diagrams show very high normalized RMS difference values between NCEP and ECMWF reanalyses of *ta* at 200 hPa in the tropics, and low values outside the tropics. Indeed, zonal mean latitude-height and 200 hPa latitude-longitude plots of *ta* (Figure 20 and Figure 21 respectively) indicate no particular problems in the tropics compared with other latitude bands. But one must keep in mind that normalization for a latitude band is by space-time variations -- including the seasonal cycle -- *within that latitude band only*. The tropics have a weak spatial gradients because of small Coriolis

effects. Thus the denominator in the normalized RMS temperature difference quotient is unusually small in the tropics. The same effect evidently occurs to some extent for outgoing shortwave radiation (*rsut*) and for specific humidity (*hus*) at 300 hPa.

Finally, we take brief note of the interannual component of the RMS difference between alternate data sets (Figure 4). Recall that this component is renormalized to interannual space-time variations, so that although it makes up only a small part of the total RMS difference, the (re)normalized differences shown in Figure 4 are often larger than the normalized totals shown in Figure 1. Figure 4 thus indicates that for sea level pressure (*psl*), geopotential height (*zg*) at 500 hPa and winds (*ua* and *va*) at 200 hPa, the difference between competing reanalyses of interannual variations is smaller than the interannual variations themselves. This condition -- evidently necessary for the data to be a reliable source for the study of interannual variations -- is not met for most of the other fields shown in Figure 4. Notable in this regard is surface air temperature (*tas*) for the Southern Hemisphere winter season. For this variable, the interannual component of RMS difference between NCEP reanalysis and the IPCC / Jones data is larger than interannual variations of the Jones data itself.

Conclusions

Rigorous "error bars" on climate observations are difficult to obtain. In their absence, the best available handle on observational uncertainty is often a simple comparison of alternative data sets. This method has obvious problems if competing data sets are insufficiently independent or of sharply differing quality. Increasing availability of observations, however, now allows most atmospheric variables to be at least crudely assessed with this technique. Different model-based reanalyses of meteorological data provide the best choices of competing observations for most variables, but in some cases more directly obtained observations (e.g., from in situ thermometers or from satellites) are preferable.

A portrait diagram displaying RMS differences between competing data sets is a useful way of compacting the resulting large amount of information. With such diagrams, we have examined 25 (single level) variables over 3 latitude bands and 4 seasons, considering both total RMS difference and its components: global mean difference (or "bias"), seasonal cycle climatology component, and interannual variability. In many ways our results confirm and quantify conventional wisdom about observational uncertainties. For example, the portrait diagrams show relatively small differences between reanalysis data sets for upper tropospheric zonal wind and relatively large differences for upper tropospheric humidity. In a few cases our results appear to contradict conventional wisdom. For tropical upper tropospheric temperature, where the portrait diagrams suggest that differences between ECMWF and NCEP reanalyses are large, the implication seems to be a misleading consequence of normalization to natural climate variations. For precipitation, however, the diagrams may reveal a more genuine surprise. Their implication of good agreement between CMAP ("Xie-Arkin") and NCEP reanalysis is consistent with other recent work (Kistler et al. 2001) and suggests that modern observational systems are resolving long-standing problems concerning precipitation. A new random error estimate for the CMAP data supports this conclusion, showing errors < 20% in the more populated areas of the globe (P. Xie, personal communication).

We have attempted to synthesize a complex and diverse collection of climatological data. The purpose of this study was to assist the overall evaluation of climate model simulations, and indeed it already has. At the same time, we recognize that evaluation of observational uncertainty is an extremely difficult task, with many experts dedicated to improving a single field or a small group of related fields. Often the identification of differences between data sources is only the beginning of an investigation. For example, as Figure 22 illustrates, a recent comprehensive report on air-sea fluxes (Taylor 2000) finds that climatological ocean surface latent heat flux (*hfls*) values can differ between reanalysis and in-situ measurements by $> 35 \text{ W m}^{-2}$. Ongoing research aims to determine which data source is more correct.

"We do not have a climate observing system at present, but rather we rely on an eclectic mix of observations, mostly taken for other purposes" (Trenberth 2002). As long as this situation continues, intercomparison of climate data sets can provide a crude but useful measure of observational uncertainty.

References

- Barkstrom, B.R., E. Harrison, G. Smith, R. Green, J. Kibler, R. Cess, and ERAB Science Team, 1989: Earth Radiation Budget Experiment (ERBE) archival and April 1985 results. *Bull. Amer. Meteor. Soc.*, **70**, 1254-1262.
- Fiorino, M., 1995: AMIP overview - Analysis of the base state, pp. 9-11 in *Proceedings of the First International AMIP Conference*, Report No. WCRP-92, World Meteorological Organization Technical Document No. 732, 532 pp. + viii.
- Fiorino, M., 2002: *PODS - The PCMDI Observational Data Base*, PCMDI Report No. XX, in preparation.
- Gibson, J.K., P. Kallberg, S. Uppala, A. Hernandez, A. Nomura, E. Serrano, 1997: ERA Description. ECMWF Re-Analysis Project Report Series, No. 1, 72 pp. European Centre for Medium-Range Weather Forecasts, RG2 9AX, Reading, UK.
- Gruber, A., and J.S. Winston, 1978: Earth-atmosphere radiative heating based on NOAA scanning radiometer instruments. *Bull. Amer. Meteor. Soc.*, **59**, 1570-1573.
- Gruber, A., and A.F. Kruegar, 1984: The status of the NOAA outgoing longwave radiation data set. *Bull. Amer. Meteor. Soc.*, **65**, 958-962.
- Gruber, A., X. Su, M. Kanamitsu and J. Schemm, 2000: The comparison of two merged rain gauge-satellite precipitation datasets, *Bull. Amer. Meteor. Soc.*, **81**, 2631-2644.
- Jones, P.D., New, M., Parker, D.E., Martin, S. and Rigor, I.G., 1999: Surface air temperature and its changes over the past 150 years. *Reviews Geophysics* **37**, 173-199.
- Jones, P.D., Osborn, T.J., Briffa, K.R., Folland, C.K., Horton, E.B., Alexander, L.V., Parker, D.E. and Rayner, N.A., 2001: Adjusting for sampling density in grid box land and ocean surface temperature time series. *Journal of Geophysical Research* **106**, 3371-3380.
- Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deavan, M. Iredell, S. Saha, G. White, J. Woolen, Y. Zhu, A. Leetmaa, R. Reynolds, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K.C. Mo, C. Ropelewski, J. Wang, R. Jenne, and D. Joseph, 1996: The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteor. Soc.*, **77**, 437-471.
- Kistler, R., E. Kalnay, W. Collins, S. Saha, G. White, J. Woolen, M. Chelliah, W. Ebisuzaki, M. Kanamitsu, V. Kousky, H. van den Dool, R. Jenne and M. Fiorino, 2001: The NCEP-NCAR 50-year reanalysis: Monthly means, CD-ROM and documentation. *Bull. Amer. Meteor. Soc.*, **82**, 247-267.
- McAvaney, B. J., C. Covey, S. Joussaume, V. Kattsov, A. Kitoh, W. Ogana, A. J. Pitman, A. J. Weaver, R. A. Wood, Z.-C. Zhao and 63 Contributing Authors, 2001: Model evaluation, pp. 471 - 521 in *Climate Change 2001: The Scientific Basis*, J. T. Houghton, Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, X. Dai, K. Maskell and C. A. Johnson, eds., Cambridge University Press, 881 pp. + x

Rossow, W.B., L.C. Garder, P.J. Lu, and A.W. Walker, 1991: International satellite cloud climatology project (ISCCP) documentation of cloud data. WMO/TD-No. 266. World Meteorological Organization, 76 pp plus appendices.

Taylor, P., ed., 2000: *Final report of the WCRP/SCOR Working Group on Air-Sea Fluxes*, WCRP-112, WMO/TD-No. 1036, 355 pp.

Trenberth, K. E., J. M. Caron and D. P. Stepaniak, 2001: The atmospheric energy budget and implications for surface fluxes and ocean heat transports. *Climate Dynamics*, **17**, 259-276.

Trenberth, K. E., D. P. Stepaniak and J. M. Caron, 2002: Accuracy of atmospheric energy budgets, *J. Climate*, submitted.

Trenberth, K., 2002: Components of a climate observing system. Extended abstract for the 82nd Annual Meeting of the American Meteorological Society, 13th Symposium on Global Change and Climate Variations, Paper 1.1. Available on the Web at <http://www.ametsoc.org/AMS>

Xie, P. and P. Arkin, 1996: Analyses of global monthly precipitation using gauge observations, satellite estimates, and numerical model predictions. *J. Climate*, **9**, 840-858.

Xie, P. and P. Arkin, 1997: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull. Amer. Meteor. Soc.*, **78**, 2539-2558.

APPENDIX: Performance Portrait Diagrams

Performance portrait diagrams provide statistical information about test data relative to a reference observational data set. In this context, "error" means the difference between the test data and the reference data. The test data can be climate model output, in which case we are measuring model errors (with the usual caveats about the quality of climate observational data). Alternatively -- as in this study -- the test data can be a set of different or "secondary" observational data.

The comparison is presented in a tabular form, but with the numerical values of the table replaced by shades of red and blue colors. The colors indicate the relative size of the root-mean-square (RMS) errors. Smaller errors are represented by blue colors, whereas larger errors are represented by red colors. Darker hues indicate the more extreme ends of the range of values.

In this study, climate variables are analyzed by these performance portraits over a selection of seasons and geographical ranges. Four different latitude bands (extending over all longitudes) are considered: global, 90°N - 20°N, 20°N - 20°S and 20°S - 90°N. For each of these latitude bands and for each of the four seasons December - January - February, March - April - May, June - July - August and September - October - November, we define the total RMS error and its components as follows:

Normalized total root-mean-square (RMS) error

In order to express the errors in non-dimensional form, which facilitates comparisons across different fields, the errors are normalized by the standard deviation of the reference field. The normalized RMS error, E , for test field f , compared with reference field r , is defined as

$$E = \frac{1}{\sigma_r} \left[\frac{1}{N} \sum_{n=1}^N (f_n - r_n)^2 \right]^{1/2} \quad (1)$$

where the sum includes all points (covering all dimensions of interest in space and time) and σ_r is the standard deviation of the reference field. When the sum includes the longitude and/or latitude dimension, the elements are weighted by grid-cell area. In this study, both the test and reference data are mapped to a 4° x 5° latitude-longitude grid before computing the statistics.

Normalized bias and normalized pattern error components of the total RMS error

The total (or uncentered) RMS error, defined by (1), can be resolved into two components in order to isolate the differences in the centered patterns from differences in the means of the two fields. One component, the overall magnitude of the normalized "bias", arises from differences in space-time means. It is defined as the absolute value of the overall mean difference between the two fields divided by the standard deviation of the reference field:

$$\bar{E} = |\bar{f} - \bar{r}| / \sigma_r \quad (2)$$

where the overbar indicates an average over all points (again covering all dimensions of interest -- space and/or time).

The other component is the normalized centered RMS or pattern difference. It is defined as the RMS difference computed after the overall bias has been removed from each field:

$$E' = \frac{1}{\sigma_r} \left\{ \frac{1}{N} \sum_{n=1}^N [(f_n - \bar{f}) - (r_n - \bar{r})]^2 \right\}^{1/2} . \quad (3)$$

From definitions (1) - (3), and from the definitions of the overall means of f and r , it follows that the two components add quadratically to give back the normalized total error:

$$E^2 = \bar{E}^2 + E'^2 . \quad (4)$$

Climatological and interannual components of the pattern error

In this study we subdivide the RMS pattern error into two components: one representing the climatological seasonal cycle and the other representing year-to-year departures, or interannual variations, from the climatological seasonal cycle. When we do so, we normalize the climatological and interannual components by the standard deviations of the climatological and interannual components, respectively, of the reference fields, rather than by the total standard deviation σ_r of the reference fields. We do this in order to better visualize the interannual component of the RMS error. Because interannual climate variations are small compared with the space-time variations associated with the climatological seasonal cycle, the interannual component of the RMS error is typically small compared with the total (or climatological) error and with σ_r . Thus, if we had chosen to normalize the interannual RMS error by σ_r , then the normalized errors in this component would appear to be nearly zero.

Our renormalization of the climatological and interannual errors prevents them from adding (quadratically) to give back the normalized pattern error. Keeping in mind the relative smallness of interannual variations, however, we can assume that the renormalized climatological component of the RMS error is about equal to the normalized RMS pattern error and thus, by itself, adds quadratically to the normalized bias to approximately give the normalized total RMS error.

Table 1: Global Variables and Data Sources

Name	Class*	Description	Primary Observations	Secondary Observations	Time Period
clt	C	total column cloud amount [percent]	ISCCP	NCEP reanalysis	1984 - 1990
hfls	C	surface latent heat flux [W m^{-2}]	NCEP reanalysis	ECMWF reanalysis	1979 - 1993
hfss	C	surface sensible heat flux [W m^{-2}]	NCEP reanalysis	ECMWF reanalysis	1979 - 1993
hur	none	relative humidity [percent]	NCEP reanalysis	ECMWF reanalysis	1979 - 1993
hus	B	specific humidity [g kg^{-1}]	NCEP reanalysis	ECMWF reanalysis	1979 - 1993
pr	C	precipitation [mm d^{-1}]	CMAP	NCEP reanalysis	1979 - 1993
prc	C	convective precipitation [mm d^{-1}]	NCEP reanalysis	ECMWF reanalysis	1979 - 1993
psl	A	sea level pressure [hPa]	NCEP reanalysis	ECMWF reanalysis	1979 - 1993
rlut	C	outgoing longwave at TOA [W m^{-2}]	ERBE	CPC	1986 - 1988
rlutcs	C	clear sky outgoing longwave at TOA [W m^{-2}]	ERBE	NCEP reanalysis	1986 - 1988
rsut	C	upward solar at TOA [W m^{-2}]	ERBE	NCEP reanalysis	1986 - 1988
ta	A	air temperature [K]	NCEP reanalysis	ECMWF reanalysis	1979 - 1993
tas	B	surface air temperature [K]	IPCC / Jones	NCEP reanalysis	1979 - 1993
tauu	C	east-west surface wind stress [N m^{-2}]	NCEP reanalysis	ECMWF reanalysis	1979 - 1993
tauv	C	north-south surface wind stress [N m^{-2}]	NCEP reanalysis	ECMWF reanalysis	1979 - 1993
ua	A	east-west wind [m s^{-1}]	NCEP reanalysis	ECMWF reanalysis	1979 - 1993
va	A	north-south wind [m s^{-1}]	NCEP reanalysis	ECMWF reanalysis	1979 - 1993
wap	none	vertical motion [Pa s^{-1}]	NCEP reanalysis	ECMWF reanalysis	1979 - 1993
zg	A	geopotential height [m]	NCEP reanalysis	ECMWF reanalysis	1979 - 1993

*Kalnay et al. (1996) classification of reanalysis variables. A = most reliable (strongly influenced by direct observations); B = less reliable (strongly influenced by model); C = least reliable (no direct observations).

References for data sources:

CMAP: Xie and Arkin 1996, 1997
CPC: Gruber and Winston 1978, Gruber and Kruegar 1984
ECMWF reanalysis: Gibson et al. 1997
ERBE: Barkstrom et al. 1989
IPCC / Jones: Jones et al. 1999, 2001
ISCCP: Rossow et al. 1991
NCEP reanalysis: Kalnay et al. 1996, Kistler et al. 2001

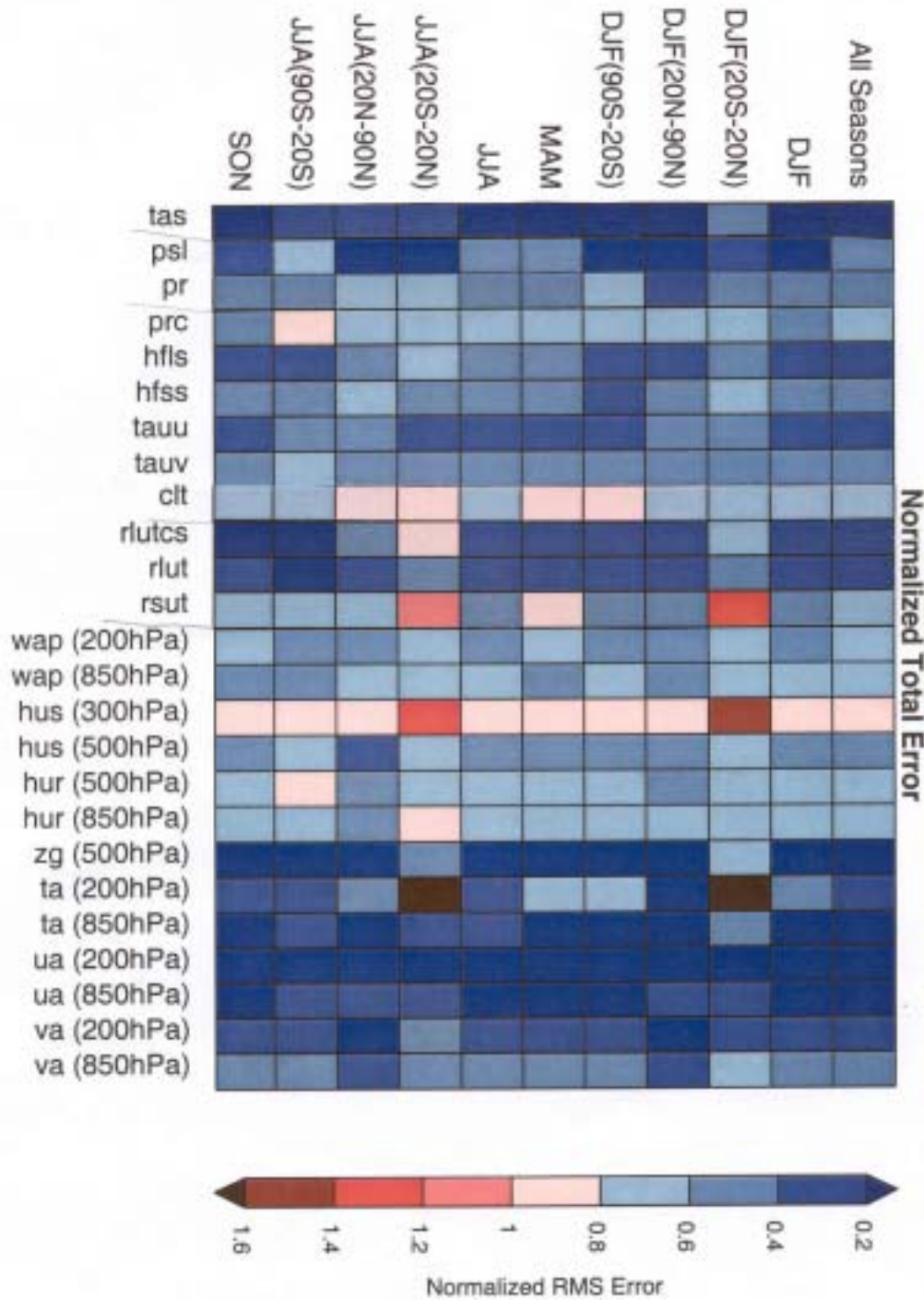


Fig. 1 (above): Total root-mean-square differences in space and time between alternate data sets. Data sets used for each of the variables (columns), together with definitions of variables, are given in Table 1. Rows divide the differences by season (December - January - February, March - April - May, June - July - August, September - October - November) and latitude band.

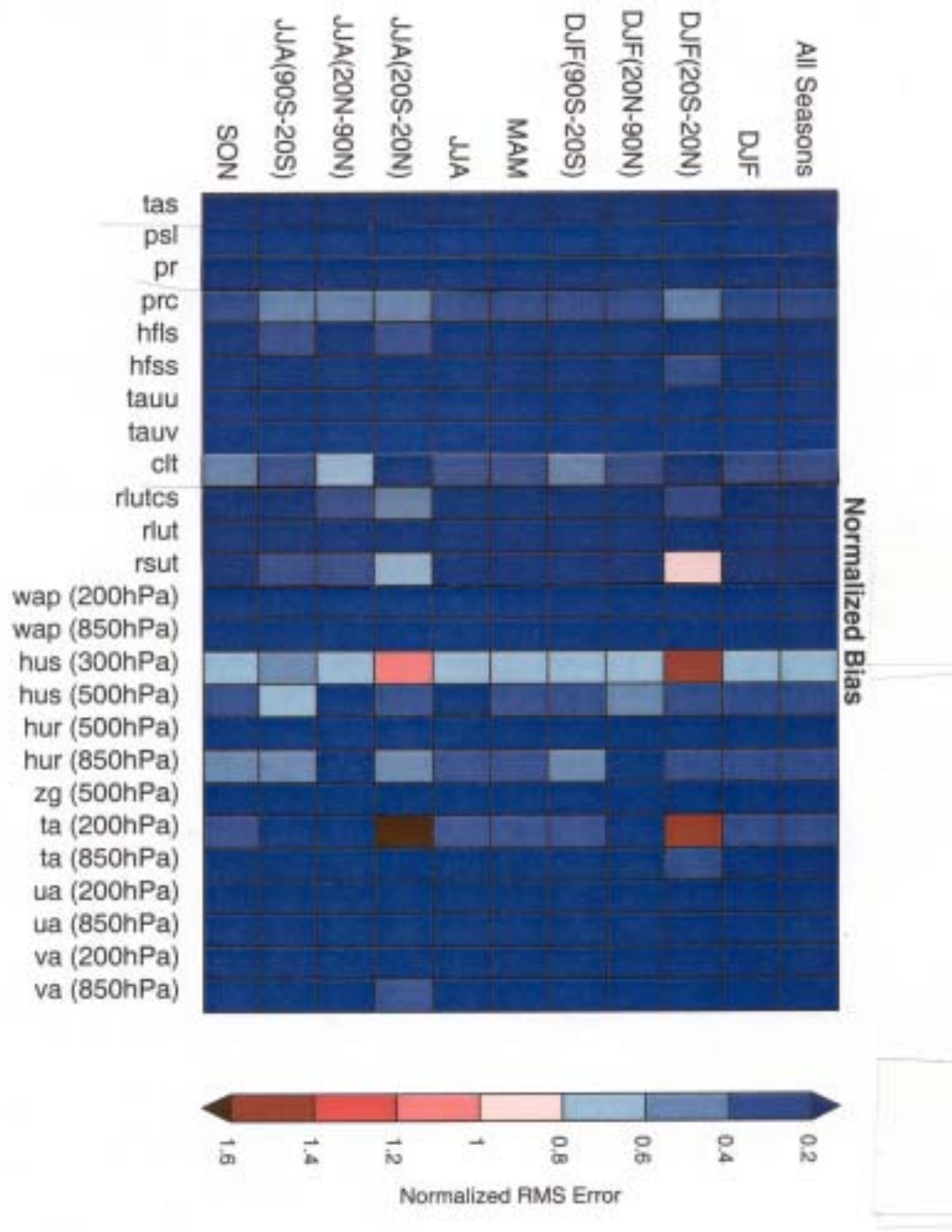


Fig. 2 (above): As in Fig. 1 for the space- and time-averaged (or "bias") component of root-mean-square differences.

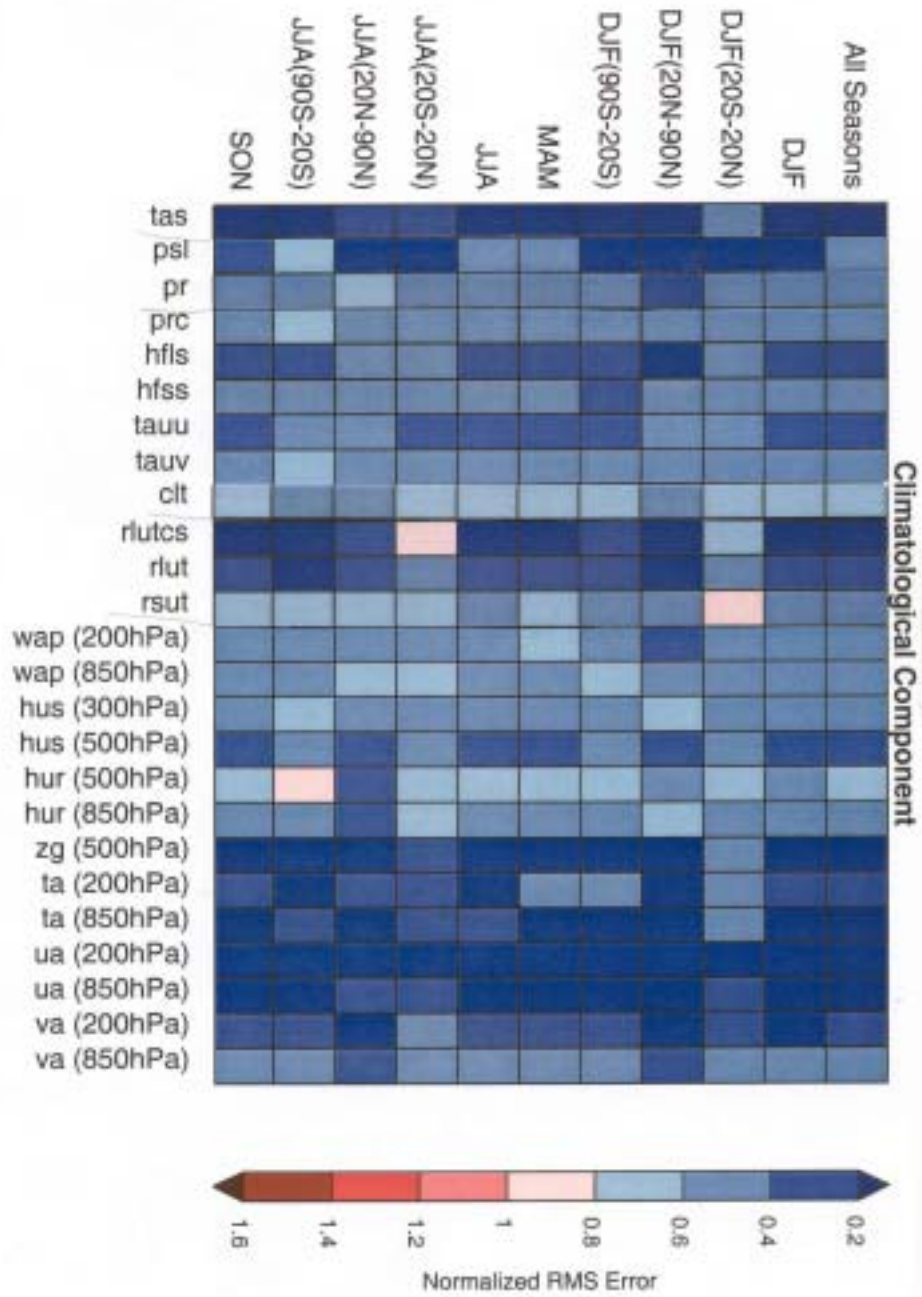


Fig. 3 (above): As in Fig. 1 for the climatological component (including seasonal cycle) of root-mean-square differences.

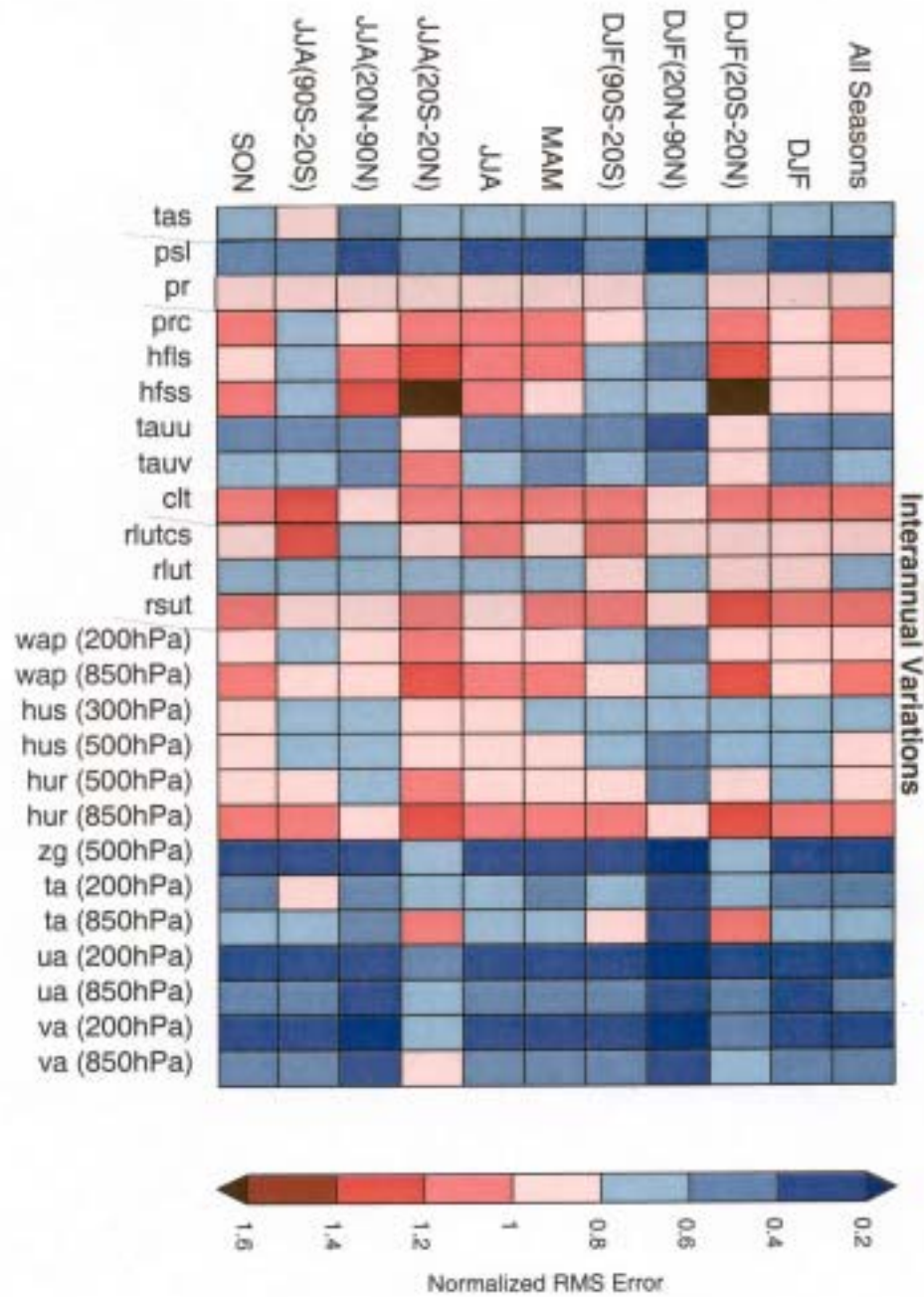


Fig. 4 (above): As in Fig. 1 for the interannual component of root-mean-square differences.

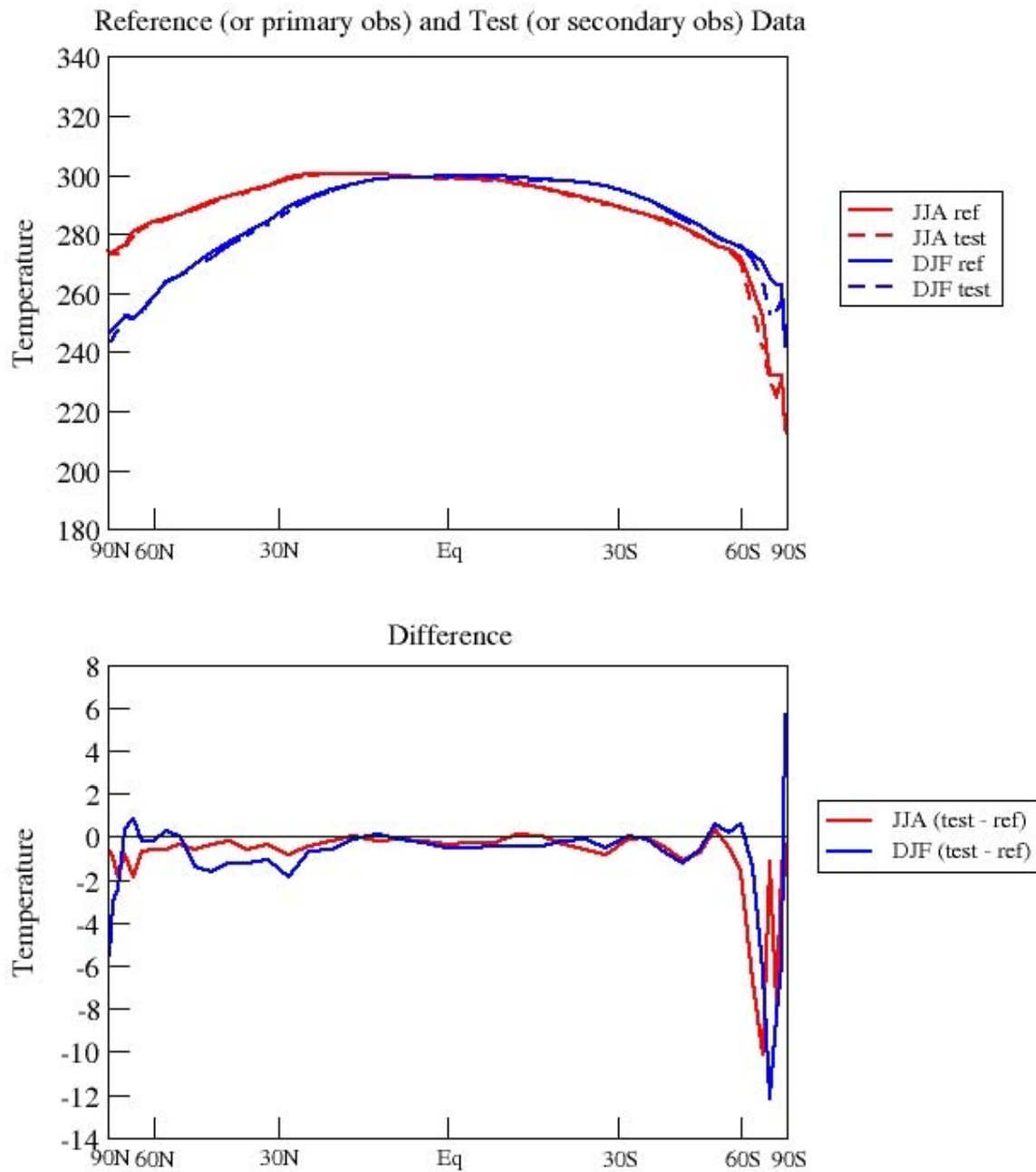


Fig. 5 (above): Climatological mean and zonal mean surface air temperature [K] for alternate data sets in December - January - February and June - July - August seasons. Primary and secondary observational data sources are identified in Table 1.

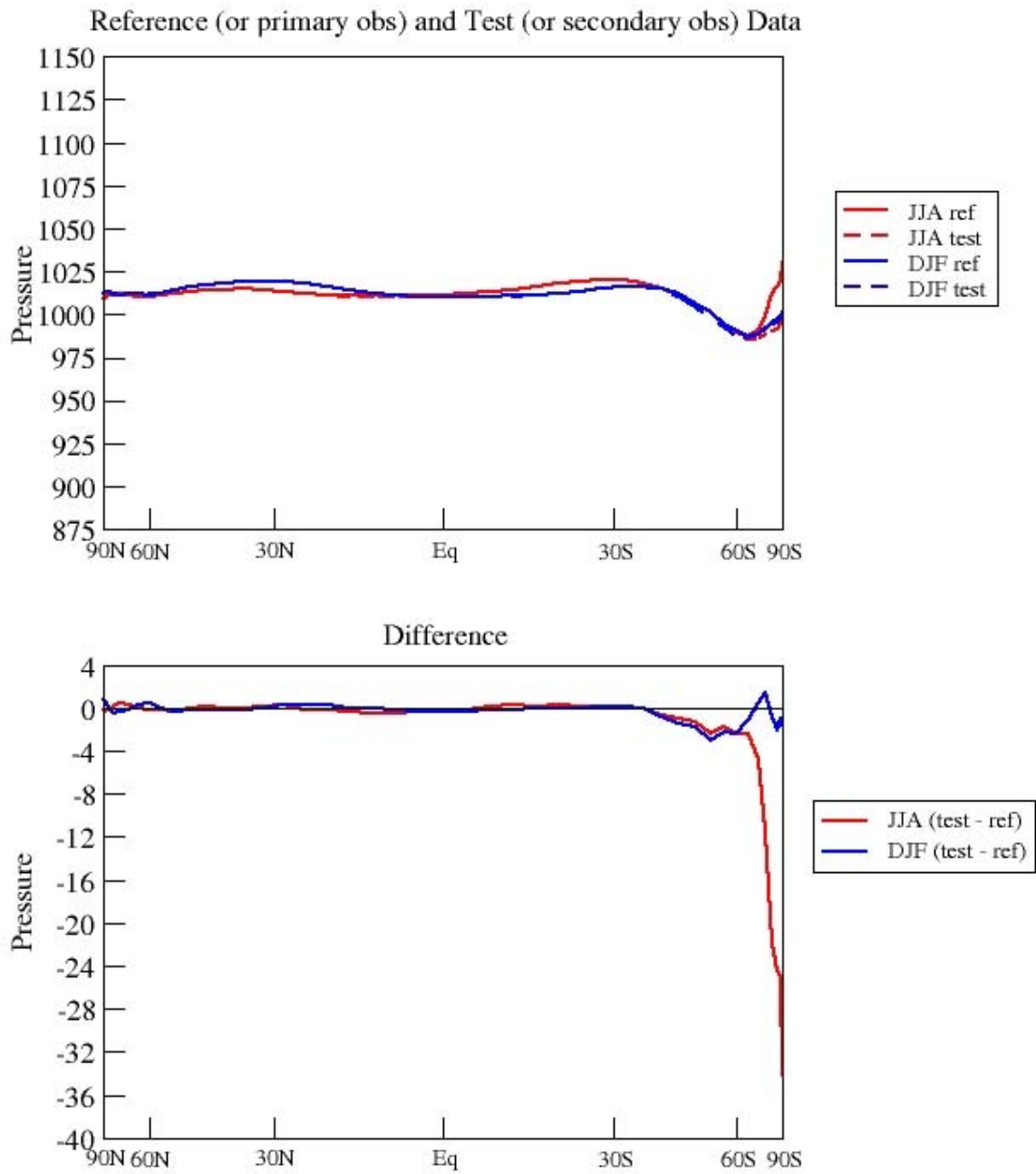


Fig. 6 (above): As in Fig. 5 for sea level pressure [hPa].

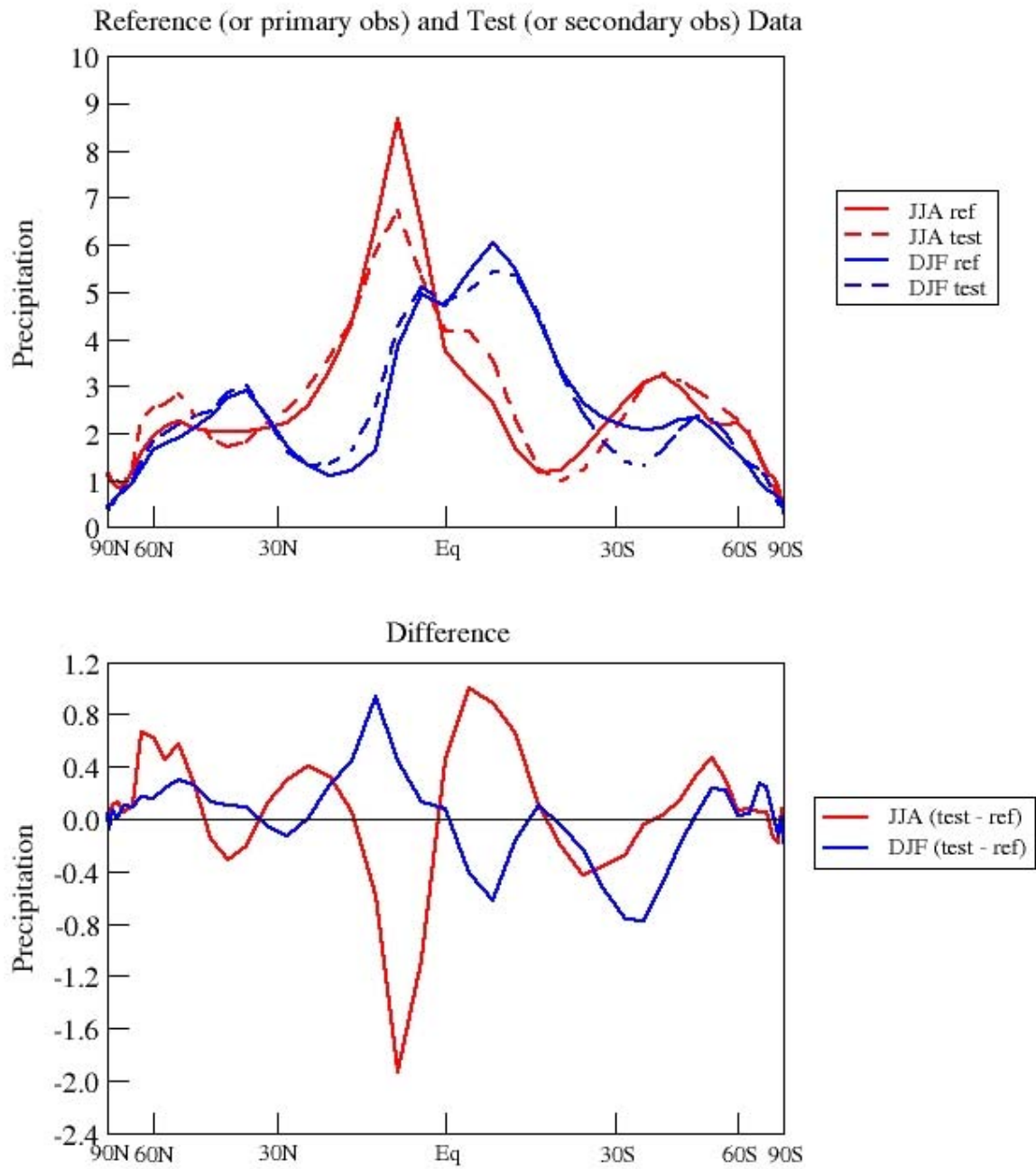


Fig. 7 (above): As in Fig. 5 for precipitation [mm d^{-1}].

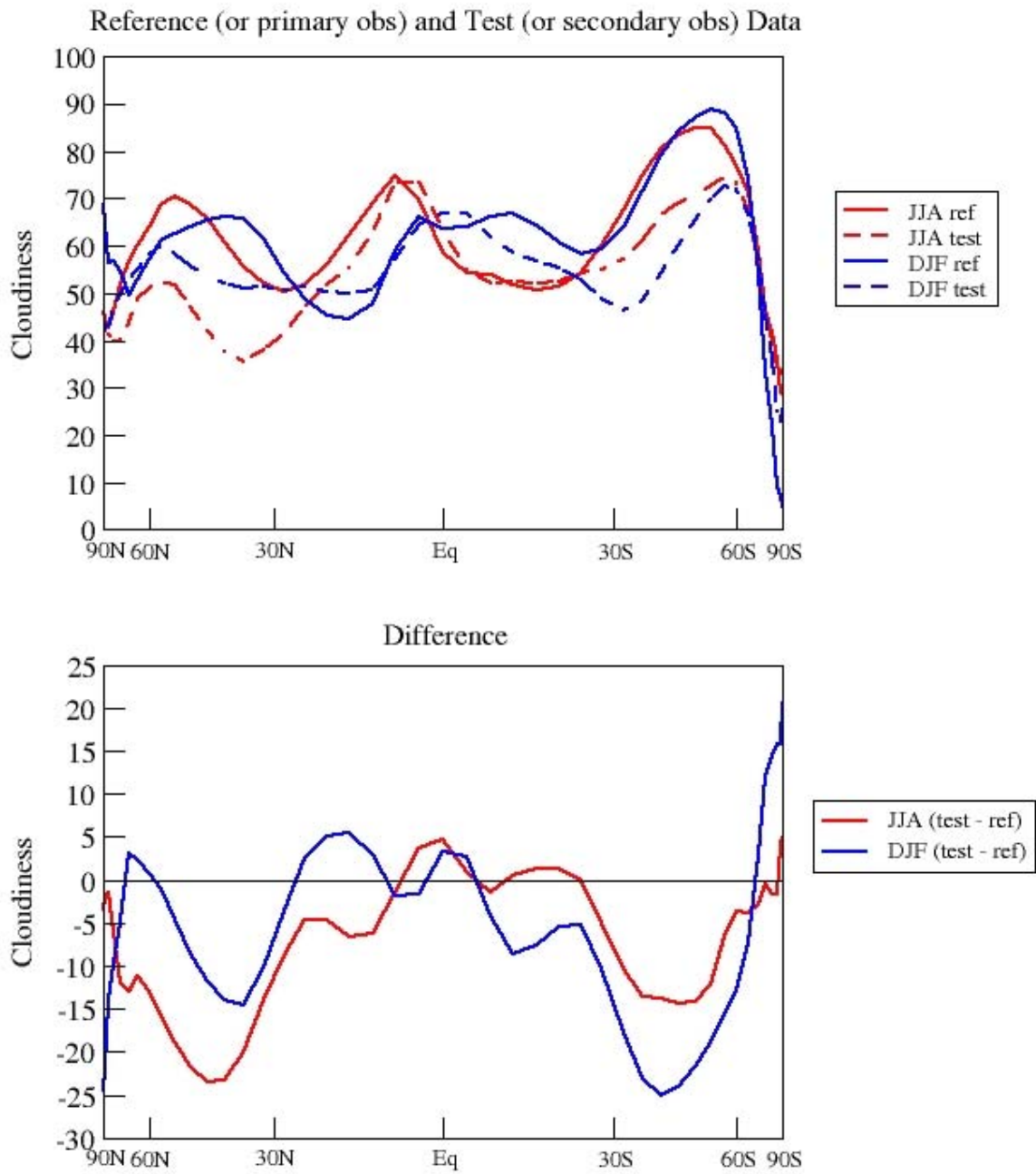


Fig. 8 (above): As in Fig. 5 for total column cloud amount [percent].

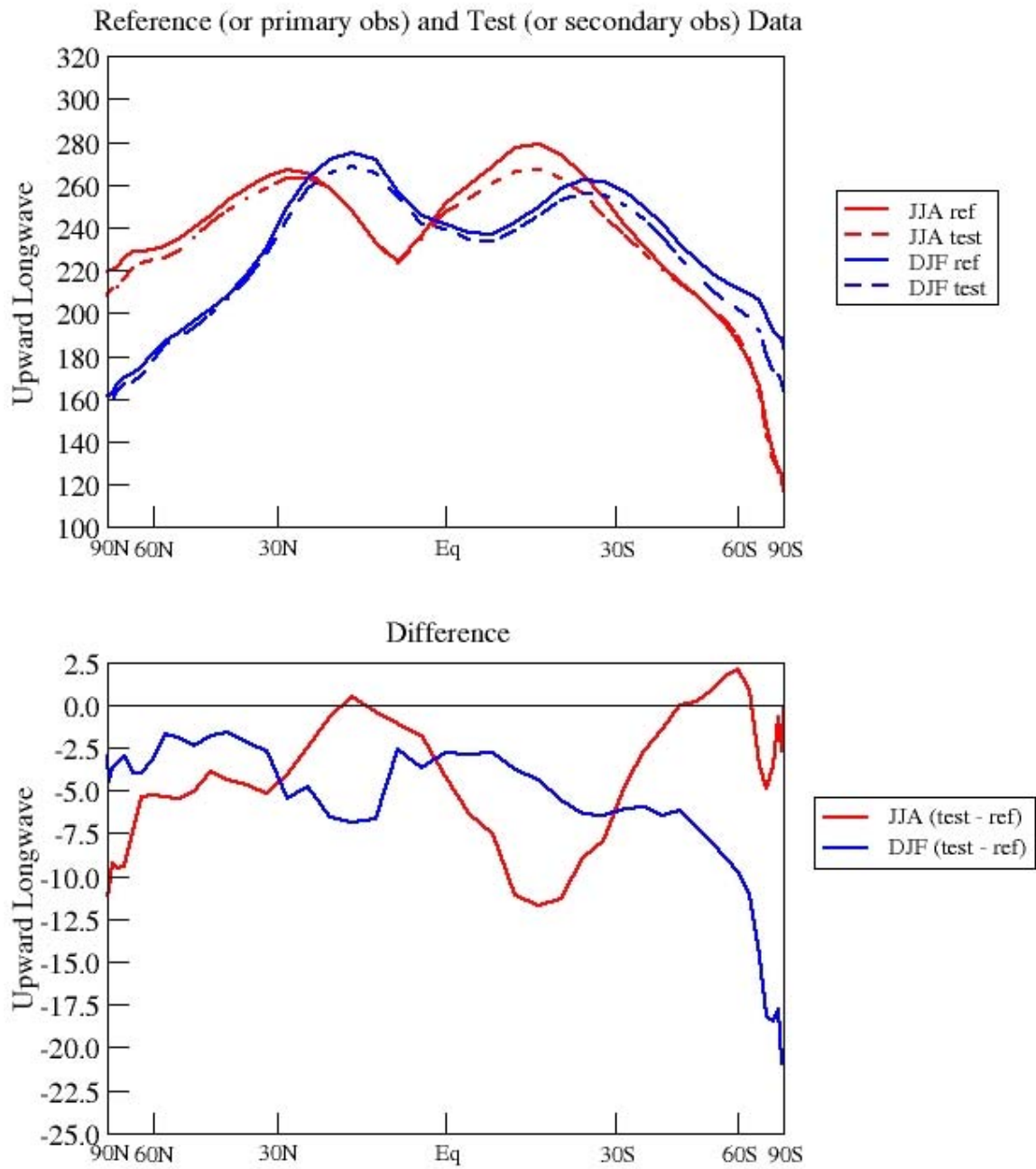


Fig. 9 (above): As in Fig. 5 for outgoing longwave radiation at the top of the atmosphere [W m^{-2}].

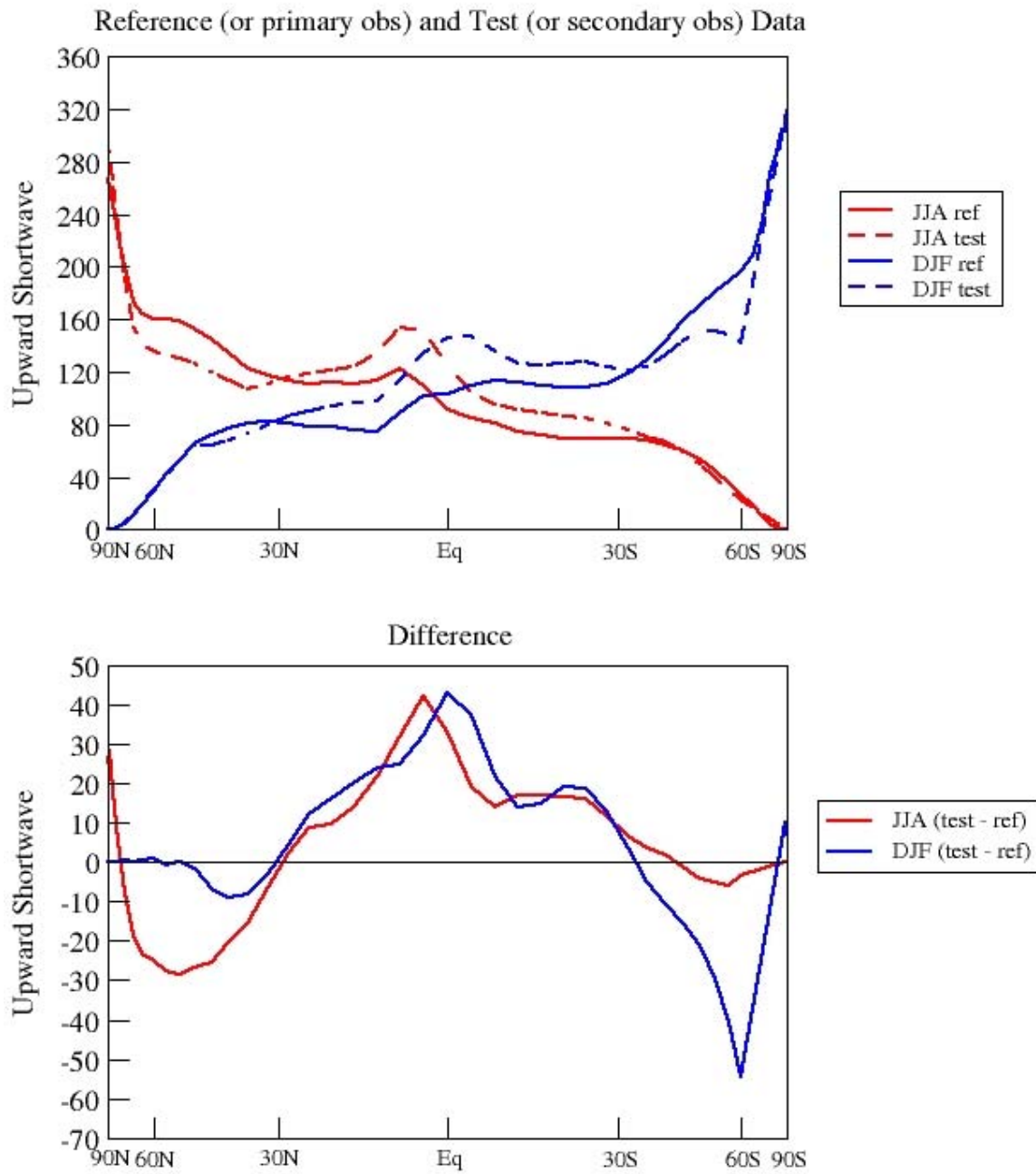


Fig. 10 (above): As in Fig. 5 for upward shortwave radiation at the top of the atmosphere [W m^{-2}].

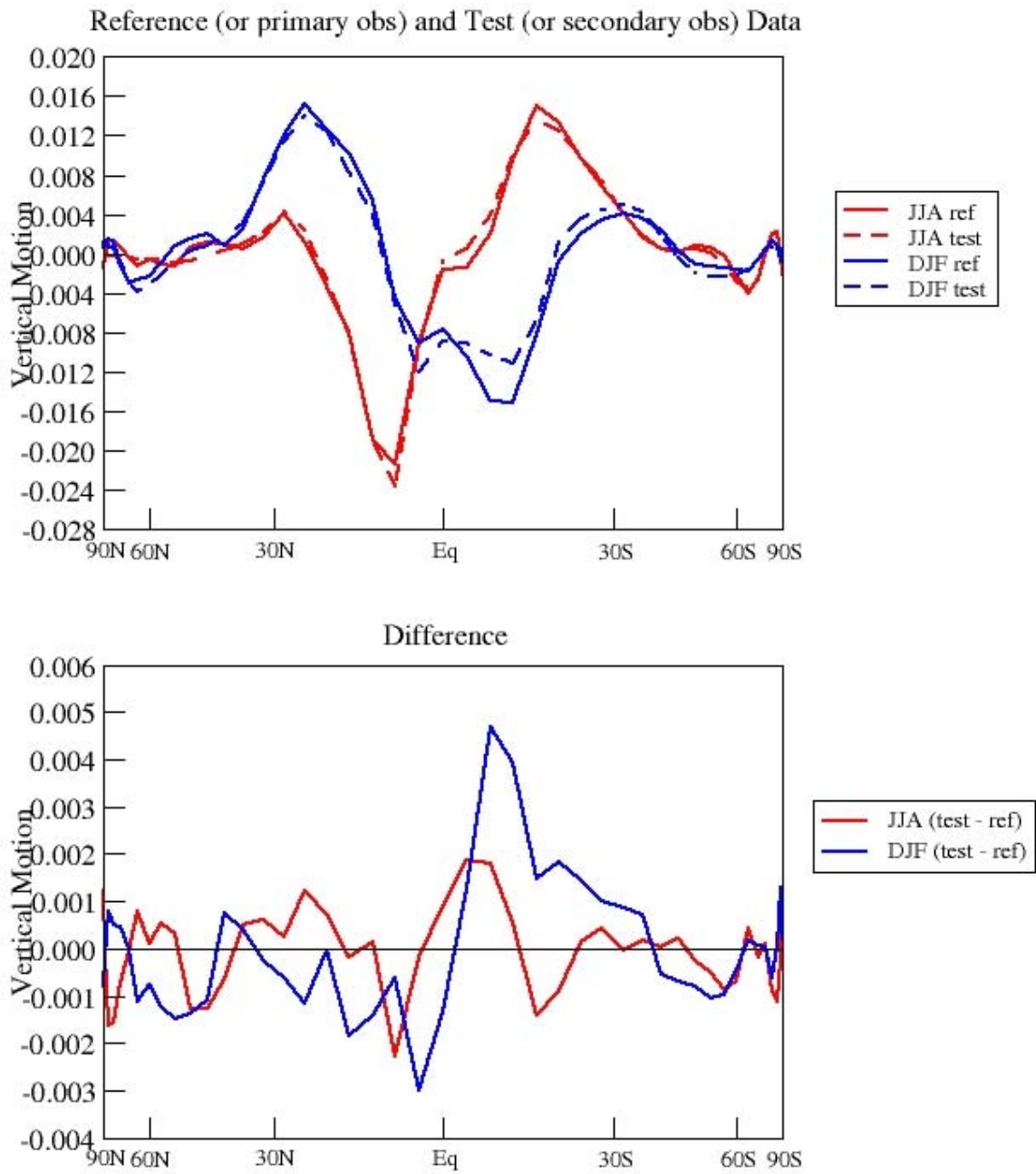


Fig. 11 (above): As in Fig. 5 for vertical motion [Pa s^{-1}] at 200 hPa.

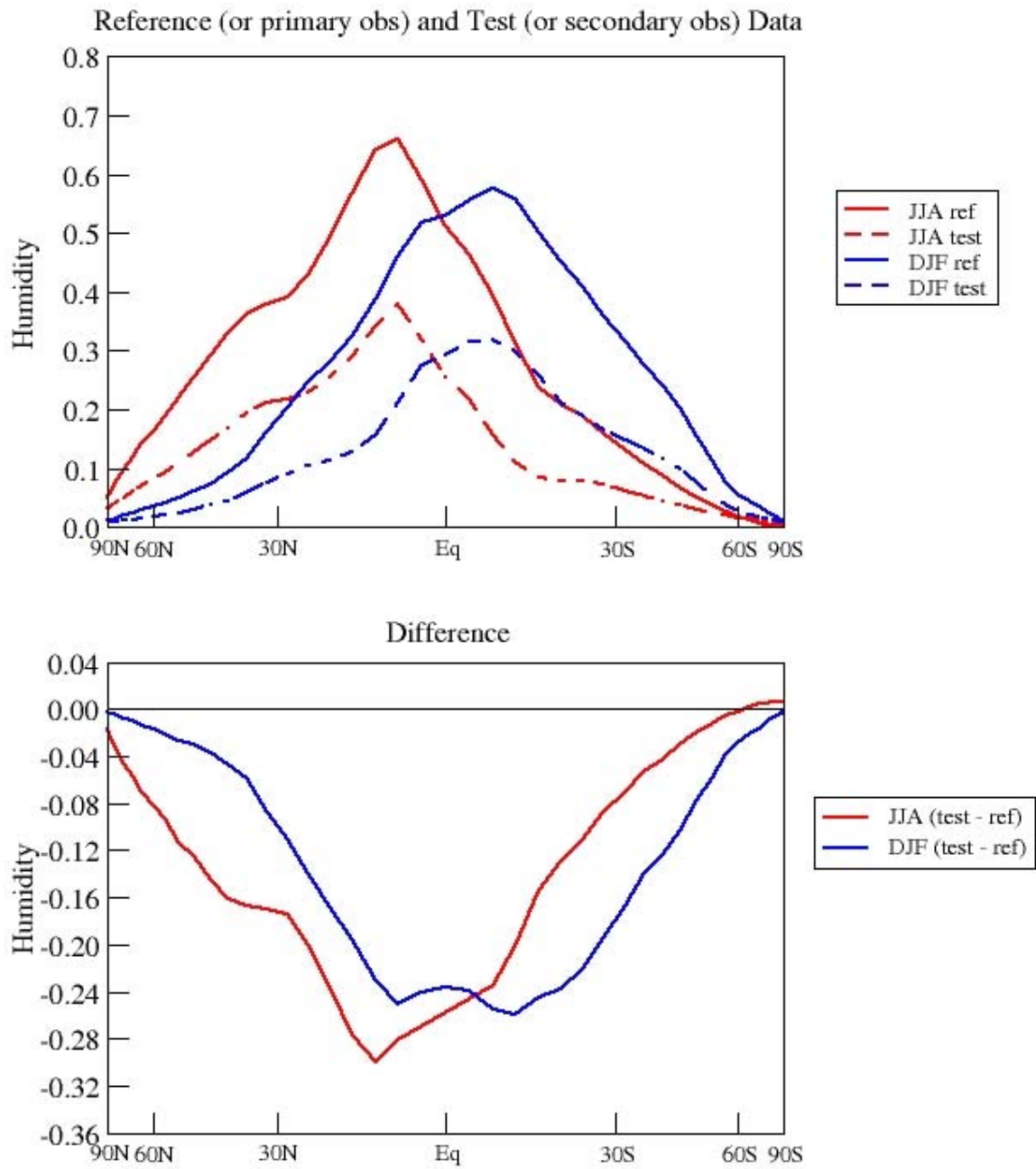


Fig. 12 (above): As in Fig. 5 for specific humidity [g kg^{-1}] at 300 hPa.

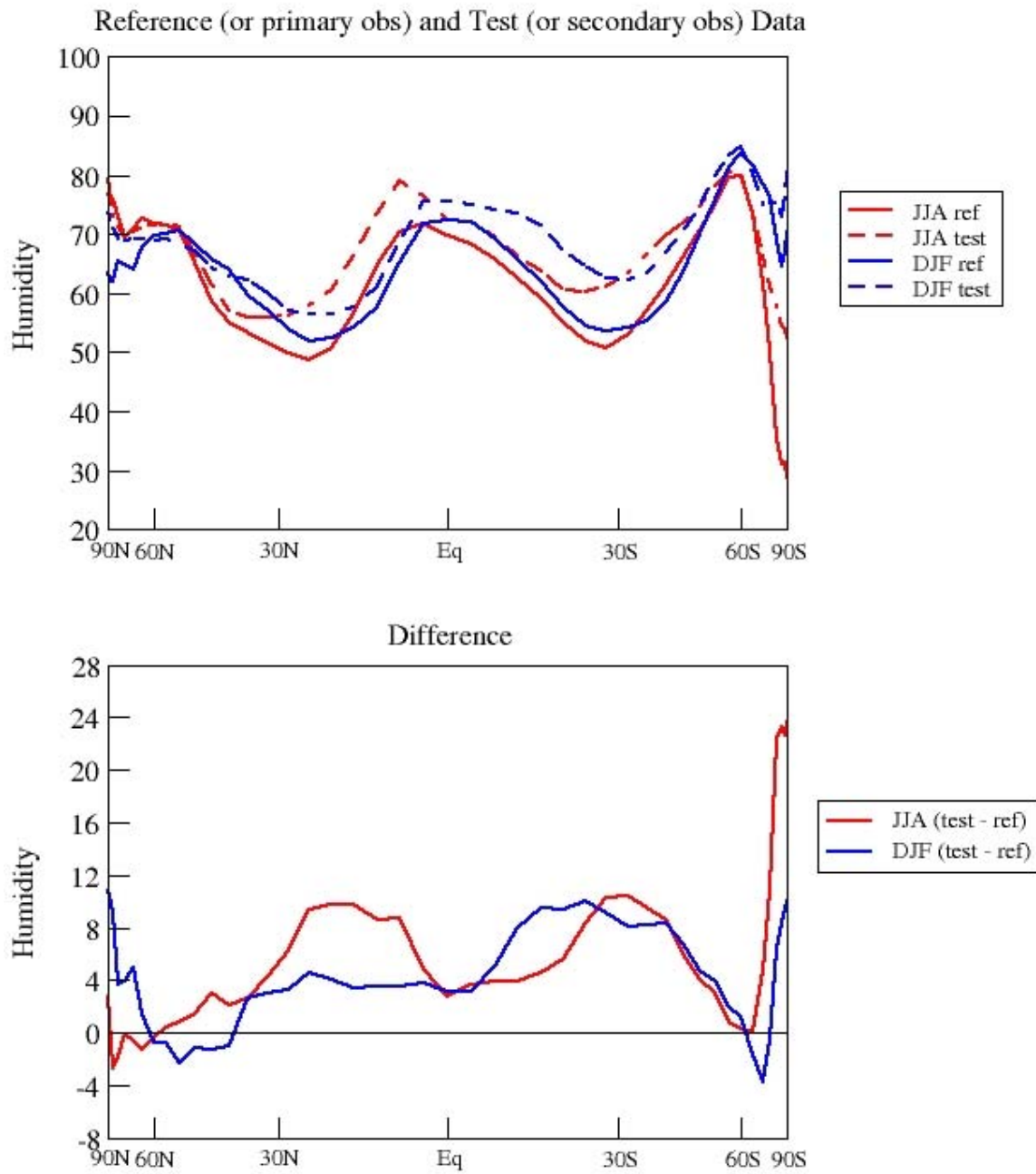


Fig. 13 (above): As in Fig. 5 for relative humidity [percent] at 850 hPa.

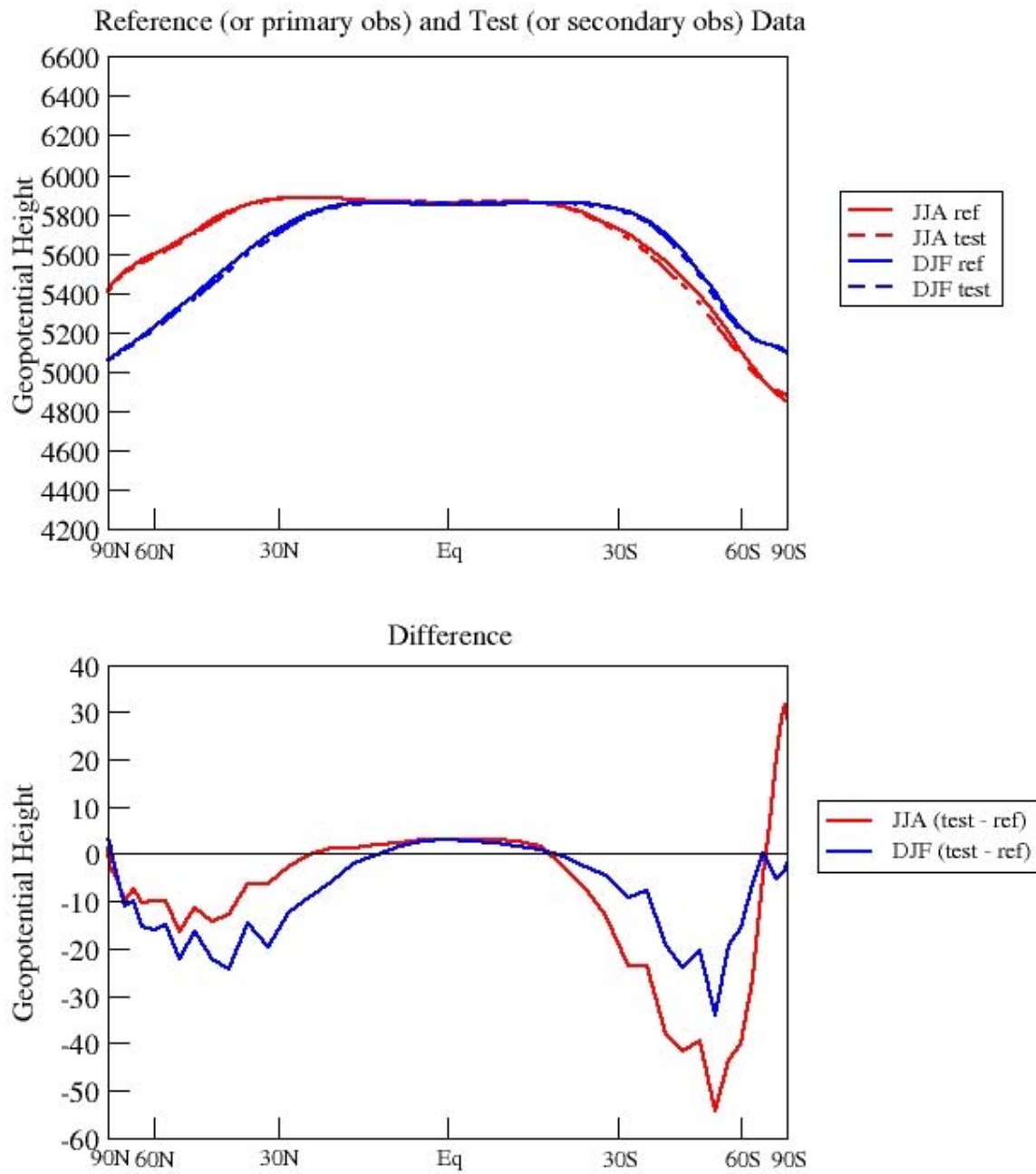


Fig. 14 (above): As in Fig. 5 for geopotential height [m] at 500 hPa.

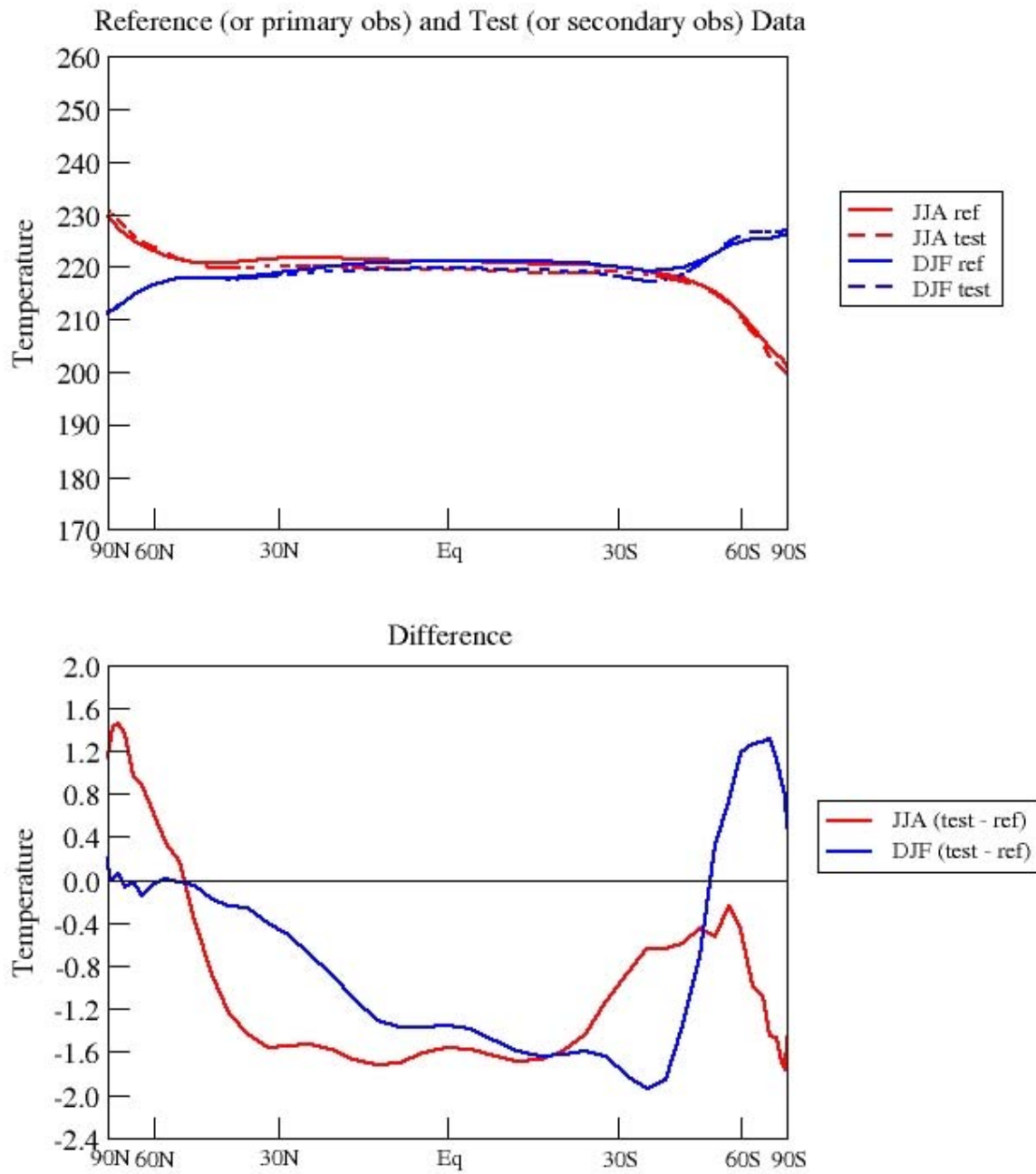


Fig. 15 (above): As in Fig. 5 for air temperature [K] at 200 hPa.

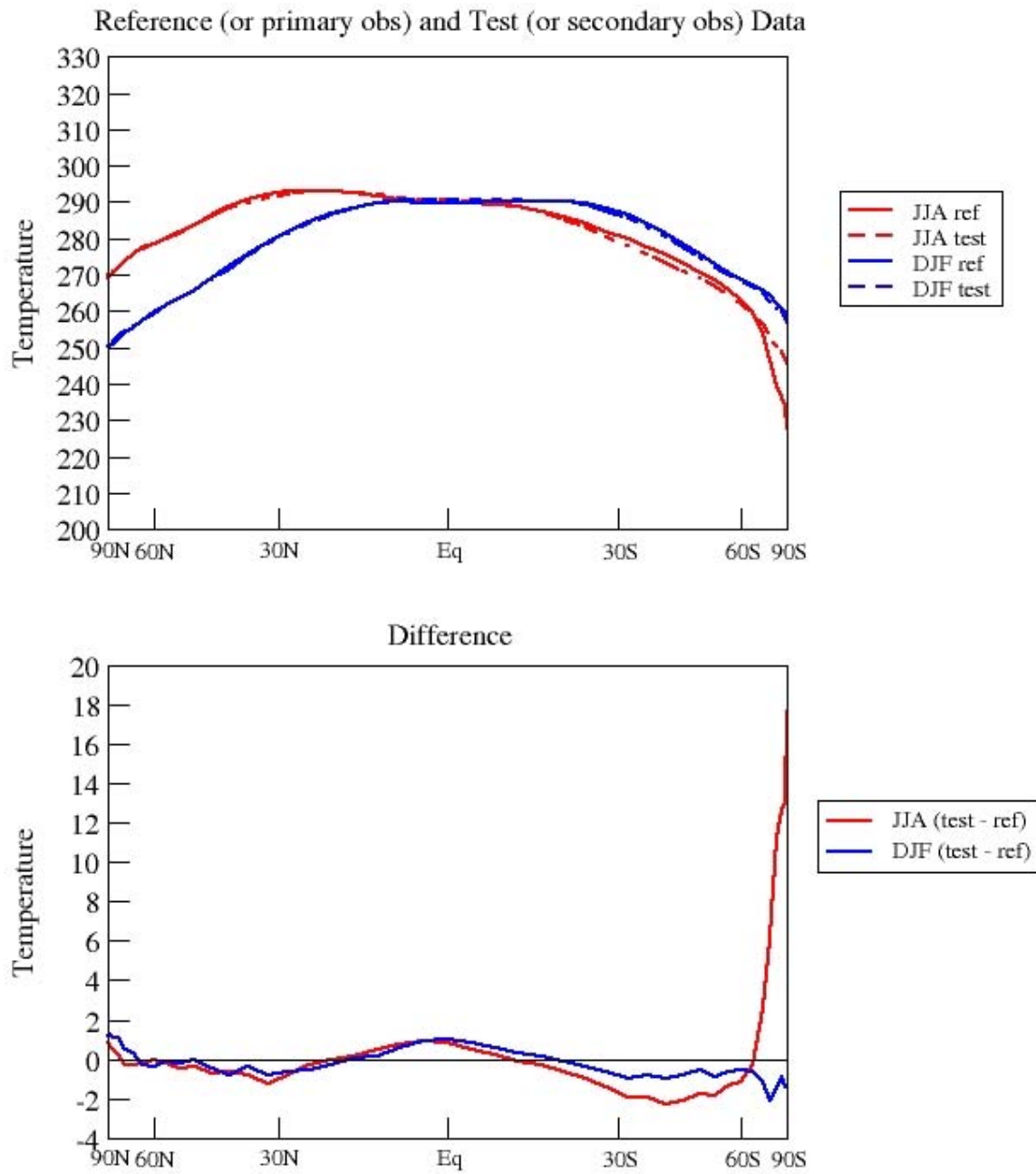


Fig. 16 (above): As in Fig. 5 for air temperature [K] at 850 hPa.

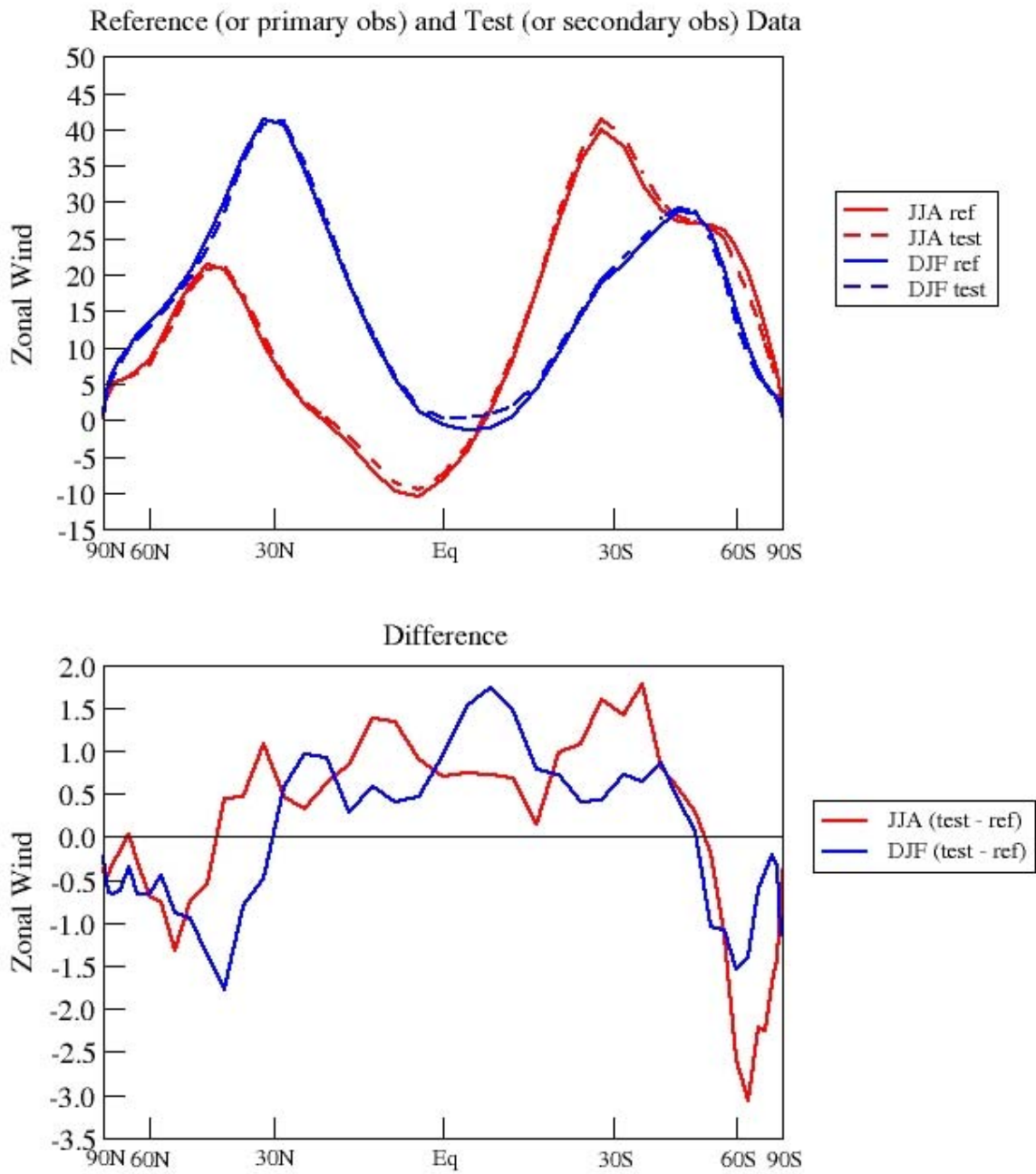


Fig. 17 (above): As in Fig. 5 for eastward wind [m s^{-1}] at 200 hPa.

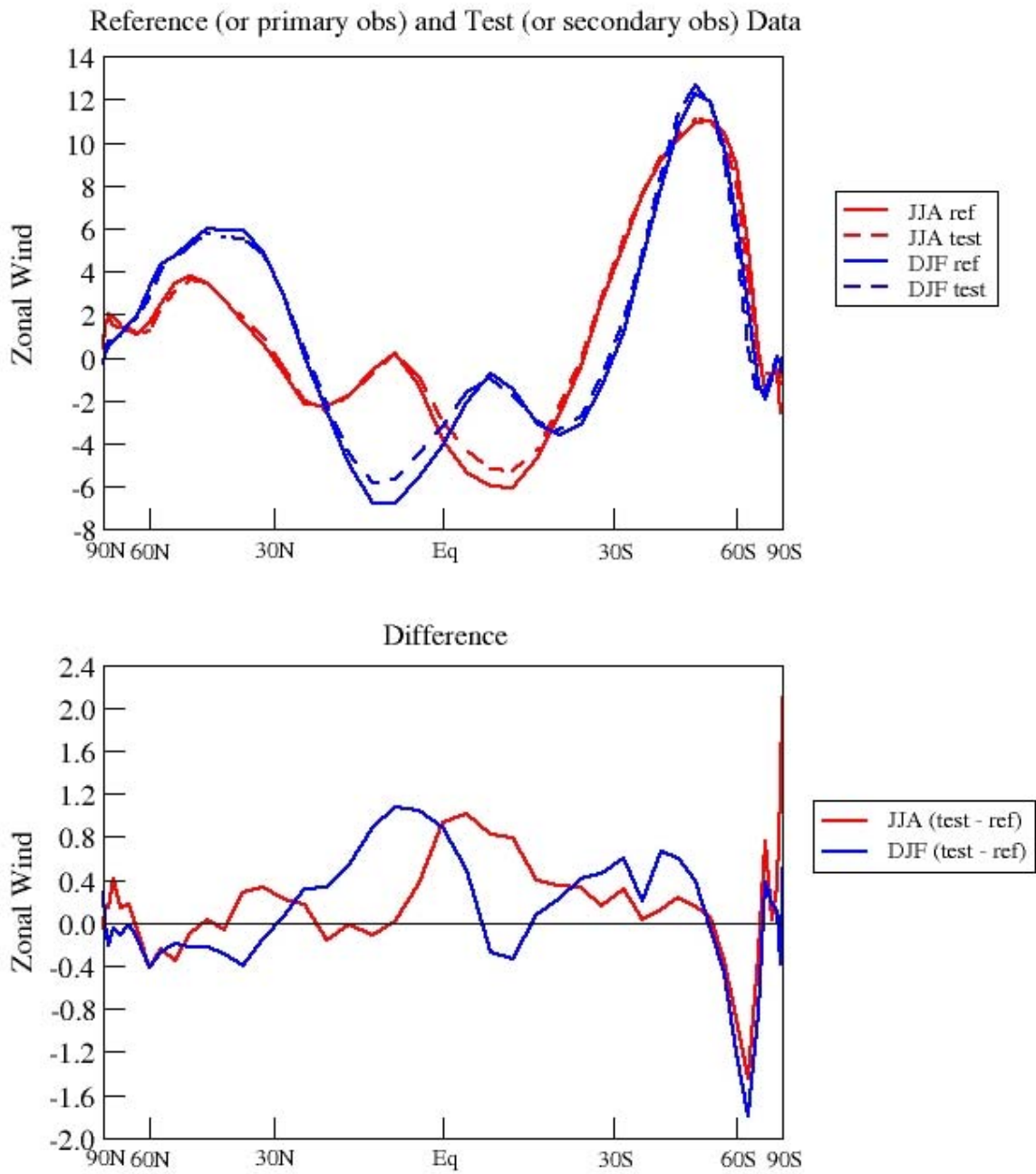


Fig. 18 (above): As in Fig. 5 for eastward wind [m s^{-1}] at 850 hPa.

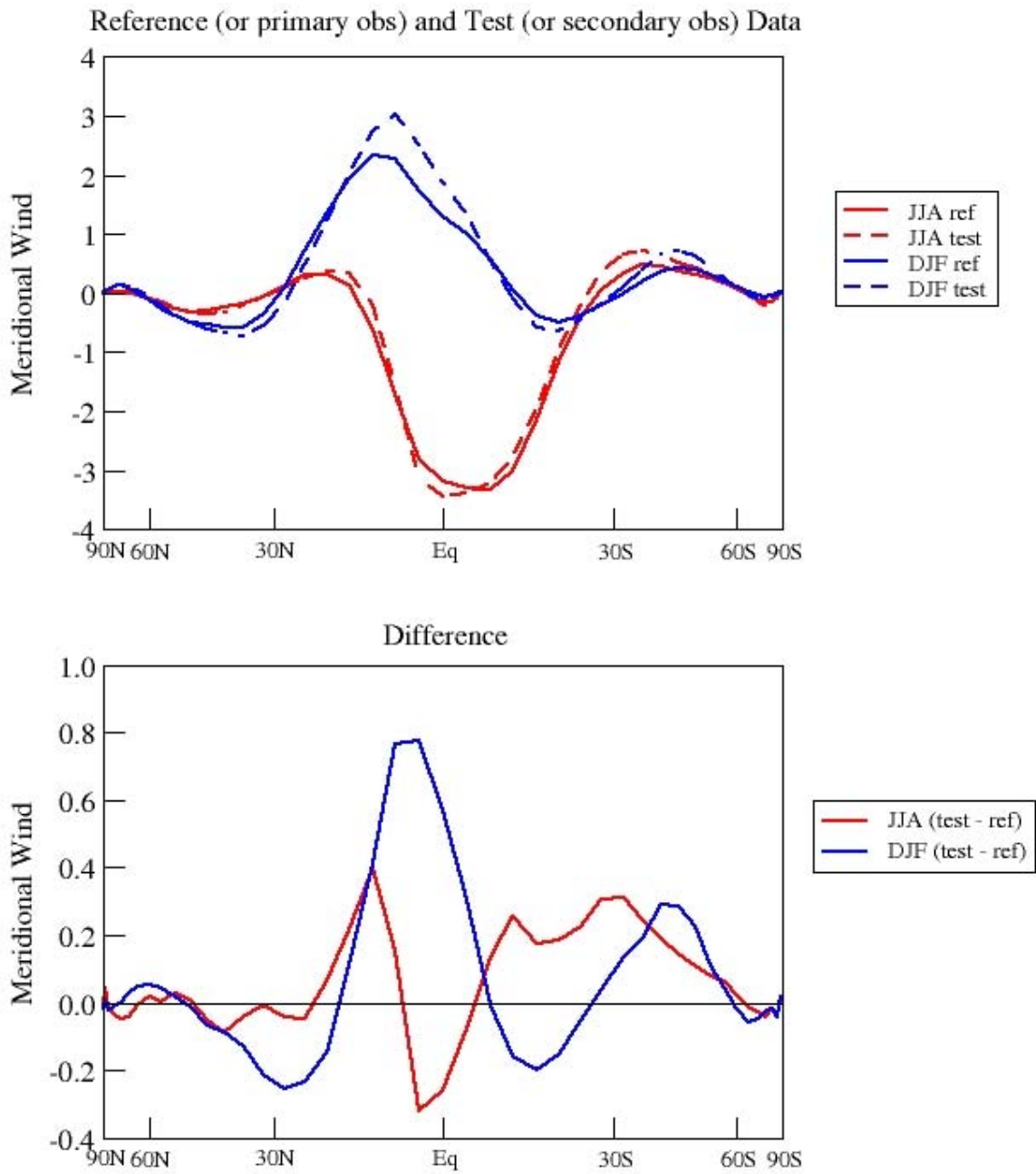


Fig. 19 (above): As in Fig. 5 for northward wind [m s^{-1}] at 200 hPa.

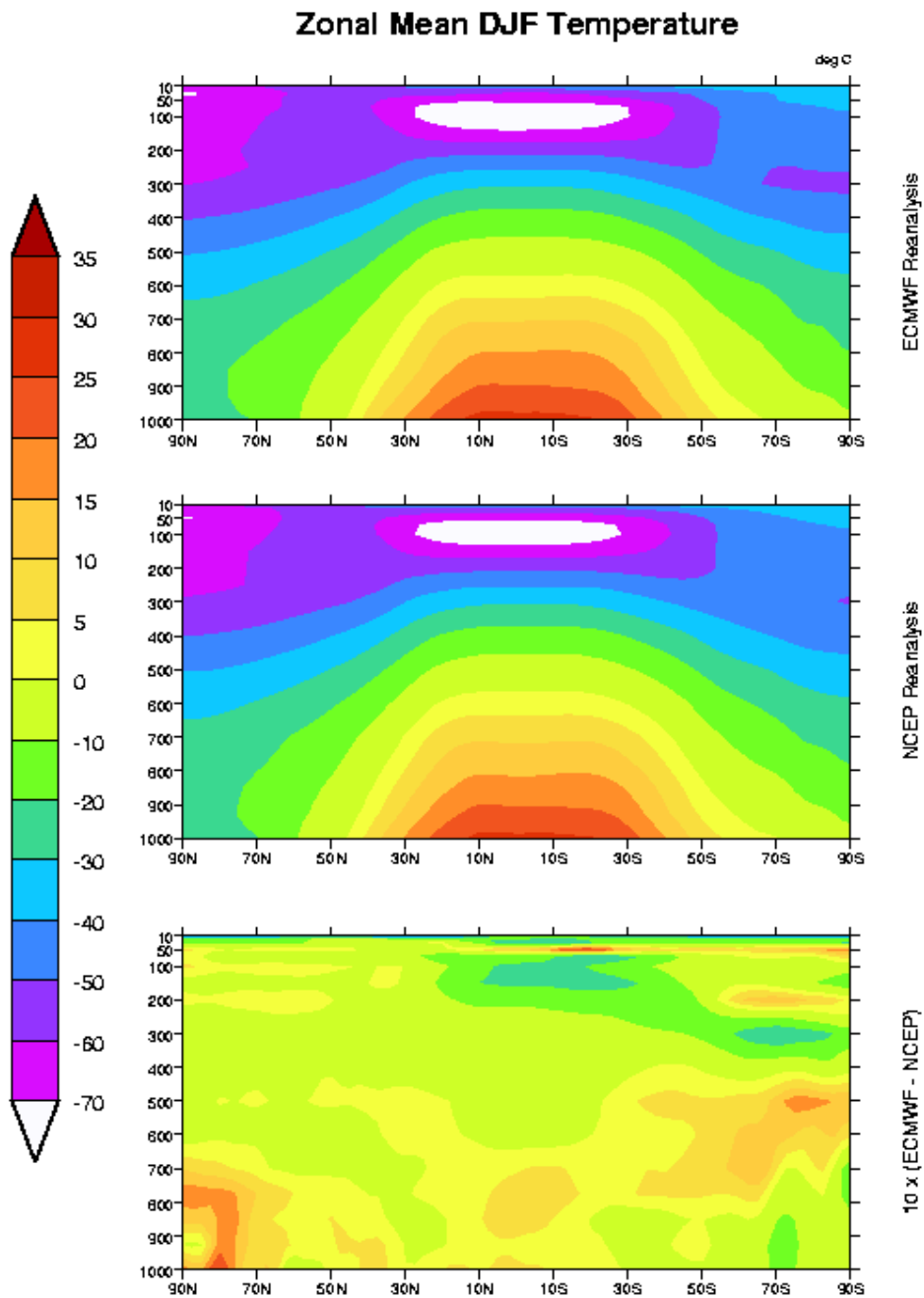


Fig. 20 (*above*): Zonal mean climatological air temperature as a function of latitude and pressure [hPa], for the December - January - February season. Shown are reanalyses from the ECMWF and NCEP and the difference between the two. The difference is multiplied by 10 to use the same color scale as the absolute values in degrees C. To emphasize features near the poles, the latitude scale is linear in contrast to the area-weighted scales in Figs. 5-19.

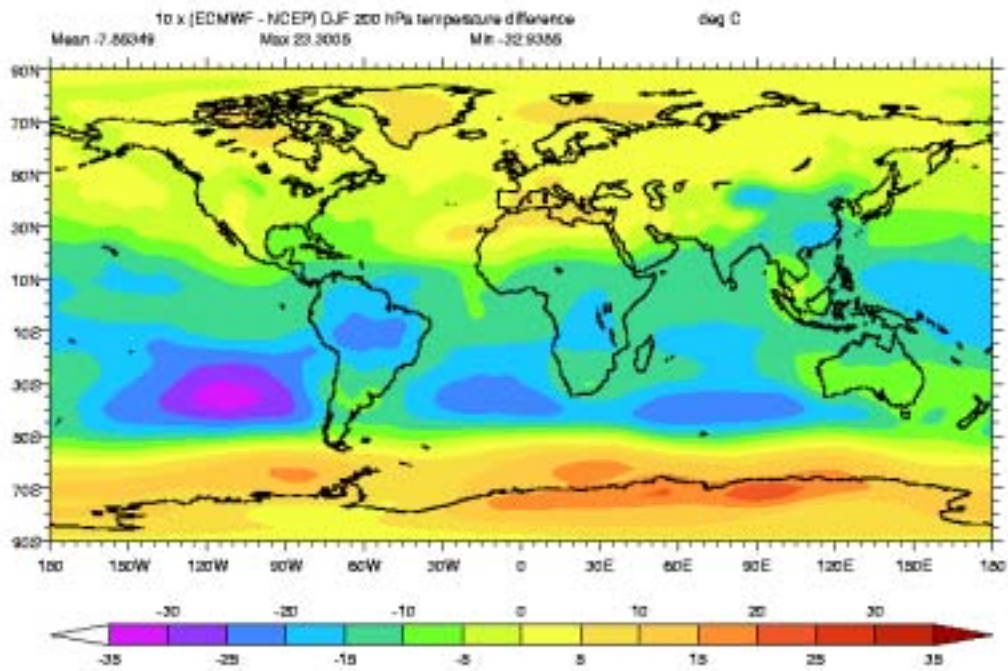


Fig. 21 (above): As in Fig. 20 for the ECMWF - NCEP difference at 200 hPa. (The difference is again multiplied by 10).

Latent Heat Flux Annual Mean Climatology (1950-1990)
NCEP Reanalysis - BOC (COADS)

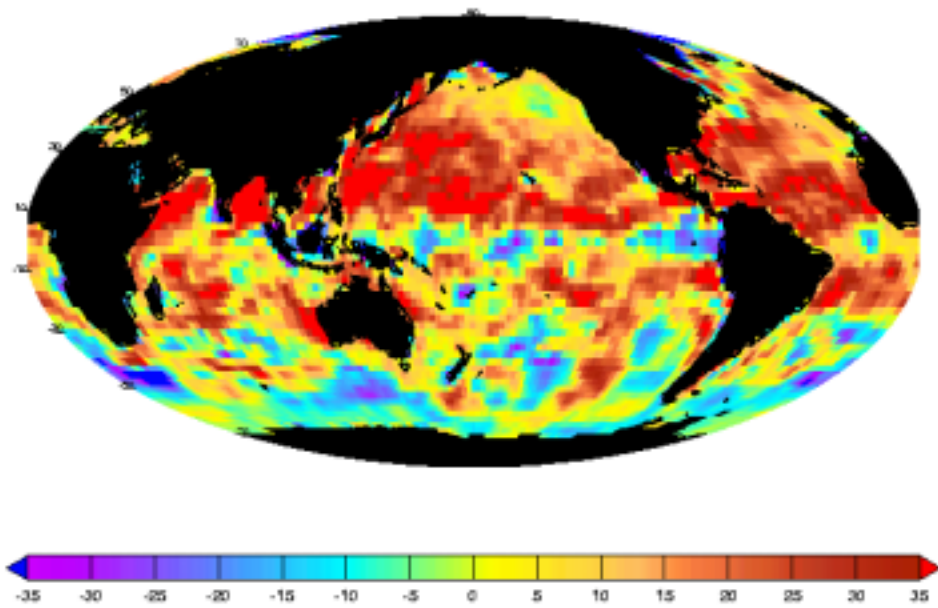


Fig. 22 (*above*): Climatological mean difference map of ocean surface latent heat flux [W m^{-2}] between NCEP reanalysis and the Southampton Oceanography Center in-situ analysis atlas. From Taylor (2002).