CMIP1 Errors and Ensembles

G.J. BOER and S.J. LAMBERT
Canadian Centre for Climate Modelling and Analysis
Meteorological Service of Canada
University of Victoria

Introduction.

Coupled global climate models (CGCMs) are widely used for studying the complex climate system, its variability, and its response to changing forcing. In practice, however, CGCMs represent a "balance of approximations" that are necessary in order to represent the working of the climate system at the finite resolutions allowed by current computer capability and the limitations of available knowledge. The steps taken to pass from the physical "laws" governing the system to a working global model introduce error and approximation. It is desirable to quantify this error, to understand its origins, and to reduce it in so far as possible. Model Intercomparison Projects (MIPs) are a "community" analysis/verification activity which seeks to: (1) document the ability of models to simulate the current climate; (2) identify common deficiencies or *systematic errors* in model results; (3) formulate hypotheses concerning the causes of model deficiencies; (4) perform experiments to clarify causes and potential solution to climate model deficiencies; and (5) to document model evolution. Coupled atmosphere/ocean models permit and exhibit behaviour absent in the separate uncoupled components and this is accompanied also by new aspects of error. Here we evaluate and intercompare basic features of the "control" or unperturbed climates of 15 coupled atmosphere/ocean models as well as that of the ensemble "mean model". The data are from the Coupled Model Intercomparison Project (CMIP) described briefly in Meehl et al. (1997).

CMIP Models and Data

The array of CMIP models differs considerably in terms of resolution and other model features as indicated in Table 1 of Lambert and Boer (2000), hereinafter LB. The atmospheric components (the AGCMs) employ spectral or finite difference numerics and, as an indication of resolution, the total number of values that specify the temperature or other variable at each timestep differs by a factor of 10 between AGCMs and by a factor of 15 between OGCMs which typically have higher resolution than their coupled AGCMs. The suite of subgrid scale physical parameterizations (radiation, convection, precipitation and cloudiness, boundary layer, etc.) in AGCMs are strong determinants of the model climate but differences among them in models are difficult to characterize simply. The OGCMs have a stronger family resemblance than do the atmospheric models since most are some version of the "modular ocean model" (MOM) model and because OGCMs have fewer subgrid scale physical parameterizations so provide less opportunity for differences in this aspect. Phillips (1998) gives basic information on the parameterizations employed in the models.

Finally, some models use "flux adjustment" or, equivalently, "anomaly coupling" to help ensure that the control climate is reasonable and that climatic feedback processes are operating in their normal range. Since the flux adjustments are proportional to the differences between simulated and observed values there is a compromise between the size of the flux adjustment and the size of the error accepted in the coupled models. Of course, as models improve, flux adjustment will become smaller, as will differences between observed and simulated values in non-flux-adjusted models, as the two approaches merge.

The atmospheric model data available for CMIP1 analysis, listed in Covey (1998), are interpolated to a 192x96xL19 grid and similarly for the ocean with L23 levels in the vertical.

Observation-based data are from a number of standard sources (see LB) and are also interpolated to the common grid.

Second order difference statistics

Observation-based quantities X and simulated values Y can each be written conceptually as $X = \overline{X} + X_f + X'$, the sum of a long-term climate mean $\overline{X}$, a function of space but not of time, other forced components X_f, which may be functions of both space and time, and the remaining random natural variability of the non-linear system where $\overline{X'} \to 0$ for sufficiently long averaging times. Both the mean and any other forced components (such as the annual and diurnal cycles or forced climate change) are deterministic in that they are the physically determined response of the system to a particular forcing. The skill of a climate simulation is in correctly reproducing the forced components together with pertinent *statistics* of the natural variability as discussed in Boer and Lambert (2000) for instance.

We deal here primarily with the global spatial patterns of mean climate variables and the implicit assumption is that the long-term averages are well determined. These means are decomposed (omitting the overbars) as $X = \langle X \rangle + [X]^+ + X^*$ under spatial averaging where $\langle X \rangle$ is the global mean, $[X]$ the zonal average, and where $[X]^+ = [X] - \langle X \rangle$ represents the north-south meridional structure and $X^* = X - [X]$ the geographical pattern about the zonal mean. The strong north-south structure is a dominant feature of many climate variables while the remaining geographic pattern, although of great practical importance, accounts for less of the total variance. Differences between simulated and observed quantities are $d = Y - X = \langle d \rangle + [d]^+ + d^*$ and the global average msd is $\langle d^2 \rangle = \langle d \rangle^2 + \langle [d]^{+2} \rangle + \langle d^{*2} \rangle$. The total msd or its components can be written in the form $\langle d^2 \rangle = \langle (Y - X)^2 \rangle = \sigma_X^2 + \sigma_Y^2 - 2\sigma_X \sigma_Y r$ which relates the msd, the variances of the observed and modelled quantity and the correlation between them.

There are 15 models in the intercomparison giving 15 values of each climatological quantity. We adopt a "multi-model ensemble" approach to the analysis of simulated climate. The ensemble view assumes that each model result represents a plausible solution to the governing equations and is an independent realization of the climate. The group of CMIP results are a sample from the set of models based on current knowledge and modelling abilities. The "mean model" result, $Y_m = \{Y\}$, is obtained by averaging over the ensemble of model results indicated by braces. Y_m may be compared with observations in the same way as an individual model result and $d_m = Y_m - X = \{Y\} - X = \{d\}$ indicates how successful models are in simulating the observed climate "on average". The ensemble or intermodel variance $\sigma_m^2 = \{(Y - \{Y\})^2\} = \{(d - \{d\})^2\}$ gives the "scatter" among model results and indicates how "consistent' the models are. To the extent that model differences are random and independent, they will cancel on averaging, $d_m = \{d\} \to 0$ and $Y_m = X + \{d\} \to X$ in the idealized limit. The virtue of the "mean model" result is that even though at any particular grid point there will very likely be several model results with less error (i.e. with $d < d_m$) when all points are considered the reverse may well be the case, i.e. $\langle d_m^2 \rangle < \langle d^2 \rangle$, and the mean model msd may be smaller than that of *any* model as discussed subsequently in Figures 1 and 2.

We take σ_m to indicate the degree of convergence of model results and d_m the average error. Other ensemble statistics are being investigated in order to further classify and analyze the information contained in the ensemble. We should like to see that both of these statistics have decreased in magnitude over time with model improvement. The ratio d_m / σ_m may be used to test if the mean model differs in a statistically significant way from the observed value. If it does so, then there is clear evidence of systematic error. Certain systematic errors are a common and persistent feature of model results (e.g. Boer et al. 1992, Gates et al. 1999) and their causes are difficult to identify and to remedy.

CMIP Results

We concentrate on surface air temperature (SAT), precipitation (P), and mean sea level pressure (mslp) which are of considerable practical importance and are associated with aspects of the energy, fresh water, and momentum fluxes between atmosphere and ocean. Figure 1 displays the globally averaged msd, $\langle d^2 \rangle$, between simulated and observed quantities and the components $\langle d \rangle^2$, $\langle [d]^{+2} \rangle$, $\langle d^{*2} \rangle$ indicating differences in the global mean, the meridional structure, and the geographical pattern. Non-flux adjusted (NFA) results indicated by enclosing their identifiers in boxes. The "mean model" result is also given.

An alternative, and in some ways more revealing, view is given in Figure 2 which displays interrelated second order statistics plotted on a new diagram as discussed in Boer and Lambert (2000) (a related diagram as used by Taylor is reported in Gates et al, 1999). The statistics for the zonal structure, $[d]^+$ and the geographical pattern d^{*2} are shown separately by dots and triangles respectively. North-south structures are comparatively better represented than the geographic patterns and this is particularly true for SAT. The geographic patterns X^* remaining after subtracting out the mean and zonal averages of the fields typically have less variance than the meridional structures and are a sensitive test of the model's ability to simulate the details of the climatological distributions of the fields. The relative msds in Figure 2 and are typically larger, and the correlations smaller (actually zero in some cases), than for meridional structures. The structure of SAT is generally best represented followed by mslp and P and that the DJF season is generally better simulated than JJA (not shown). Both figures indicate that FA model results are generally closer to the observations, at least for this generation of models, than are NFA model results although some individual NFA model results are better than some FA model results. The figures also show that different models have different errors in different variables and that no model is best for all.

According to Figures 1 the "mean model" is generally the "best model" when measured in terms of the msd. This is also the case for the related statistics of Figure 2 and particularly so for the geographic pattern (indicated by the triangles) where the mean model result is visually separate from the cloud of individual model results.

LB displays a variety of atmospheric and oceanic statistics which cannot be reproduced here. Figure 3 displays basic zonally averaged distributions of mean sea level pressure and precipitation results from coupled models and from observation-based climatologies.They may be compared with similar results for several generations of models developed over several decades and shown in Boer (2000). Results show that persistent and characteristic differences remain, although some progress has been made.

Although atmospheric models have been intercompared for some time, this is not the case for coupled models although there have been some efforts in that direction (Gates et al., 1993, IPCC1995). It seems safe to say at this state of model development that oceanic results in coupled models differ more widely than do atmospheric results, and this is especially true at depth. As an example of one very basic ocean parameter we show as part of Figure 3 some results for the salinity distribution in CMIP models.

Summary

Results for coupled models may be summarized, as in LB, as follows: (1) the current generation of climate models, on average, reproduce the major features of the observed distribution of the basic climate parameters; (2) there is, nevertheless, a considerable scatter among model results and between simulated and observed values; (3) this is particularly true of oceanic variables; (4) flux adjusted models generally produce simulated climates which are in better accord with observations than do non-flux adjusted models (as, of course, is the intent of flux adjustment); (5) some non-flux adjusted model results are, nevertheless, closer to the observations than some flux adjusted model results; (6) systematic differences, i.e. differences common to most models, are seen and some of these, i.e. for precipitation and mean sea-level pressure, have been known for some time and show a slow improvement with model evolution; (7) other model differences, such as resolution, do not appear to provide a clear distinction among model results in this generation of models; (8) as is characteristic of intercomparison results, different climate variables are simulated with different levels of success by different models and no one model is “best” for all variables; and (9) there is some evidence that the “mean model” result, obtained by averaging over the ensemble of model results, provides an overall “best” comparison to observations for climatological mean fields.

References

Boer, G. J., K. Arpe, M. Blackburn, M. Deque, W. L. Gates, T. L. Hart, H. Le Treut, E. Roeckner, D. A. Sheinin, I. Simmonds, R. N. B. Smith, T. Tokioka, R. T. Wetherald and D. Williamson, 1992: Some results from an intercomparison of the climates simulated by 14 atmospheric general circulation models. *J. of Geophys. Res.*, **97**, 12771-12786

Boer, G.J., 2000: Climate model intercomparison. Chapter 3 in “Numerical modelling of the global atmosphere in the climate system”. Kluwer Academic Publishers, Dordrecht, the Netherlands.

Boer, G.J. and S.J. Lambert, 2000: Second order space-time climate difference statistics. Accepted *Clim. Dyn*.

Covey, C., 1998: CMIP1 model output. At http://www-pcmdi.llnl.gov/cmip/diagsub.html#CMIP1 model output. Program for Climate Model Diagnosis and Intercomparison (PCMDI), the Lawrence Livermore National Laboratory, Livermore, California

Gates, W.L., U. Cubash, G.A. Meehl. J.F.B. Mitchell and R.J. Stouffer, 1993: An intercomparison of selected features of the control climates simulated by coupled ocean-atmosphere general circulation models. World Climate Research Programme, WCRP-82. WMO/TD No. 574, WMO, Geneva.

Gates, W.L., et al., 1999: An overview of the results of the Atmospheric Model Intercomparison Project (AMIP), *Bull. Amer. Meteor. Soc.*, **80**, 29-56.

IPCC, 1995: Chapter 5, Climate Models - Evaluation, in *Climate Change 1995: The Science of Climate Change*. J T Houghton, L G Meira Filho, B A Callender, N Harris, A Kattenberg and K Maskell (Eds.). Cambridge University Press, UK. 572pp. (ISBN: 0-521-56433-6)

Lambert, S.J. and G.J. Boer, 2000: CMIP1 evaluation and intercomparison of coupled climate models.Accepted *Clim. Dyn*.

Meehl, G.A., G.J. Boer, C. Covey, M. Latif, and R. J. Stouffer, 1997:Intercomparison makes for a better climate model. *EOS, Trans. Amer. Geophys. Union*, **78**, 445-446, 451.

Phillips, T., 1998: Summary Documentation: CMIP I Model Features and Experimental Implementation (Version 1.0). At http://www-pcmdi.llnl.gov/modeldoc/cmip/index.html. Program for Climate Model Diagnosis and Intercomparison (PCMDI), Lawrence Livermore National Laboratory, Livermore, California

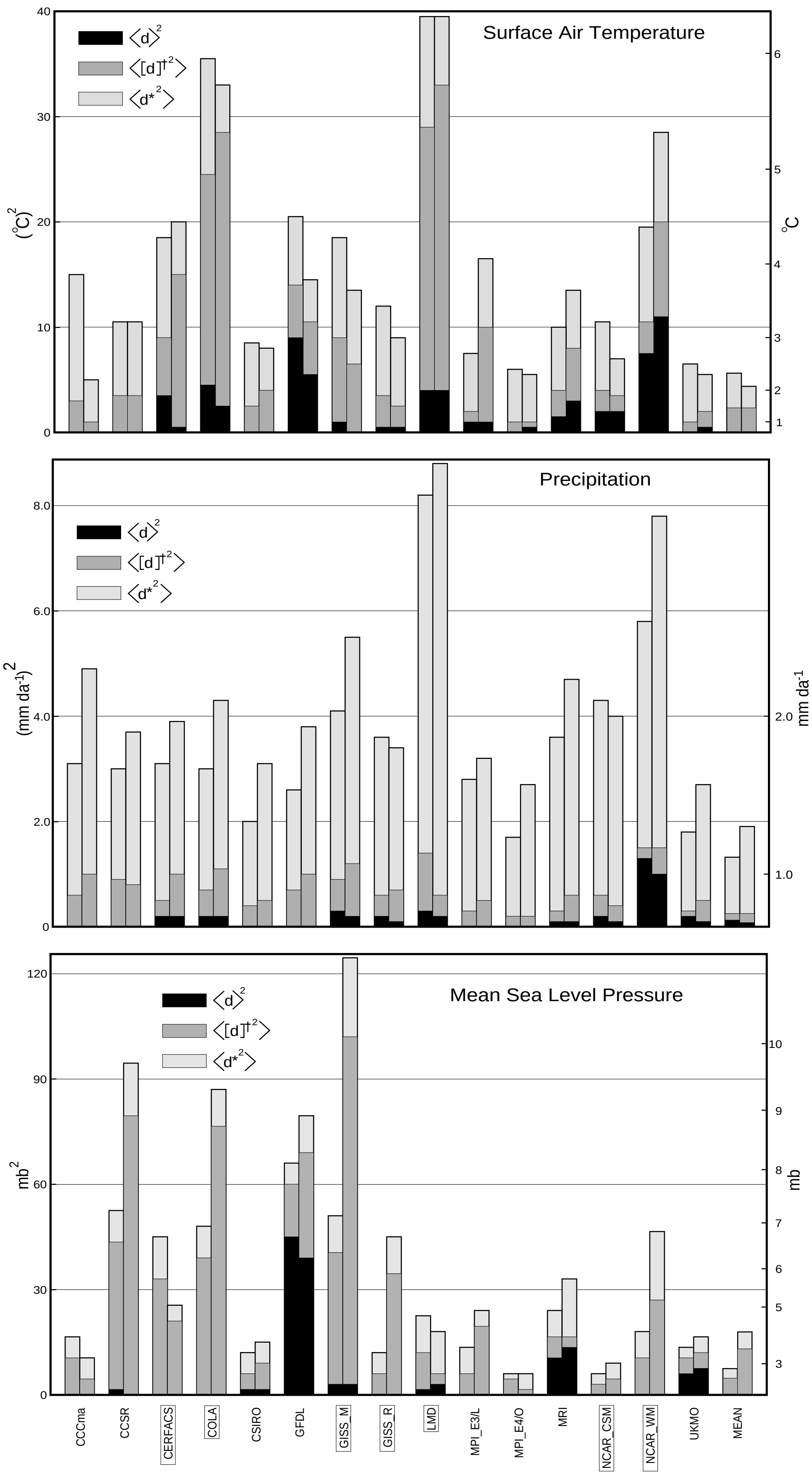

Figure 1. Mean square differences from observation-based climatology or each model and for the "mean model". DJF on left and JJA on right.

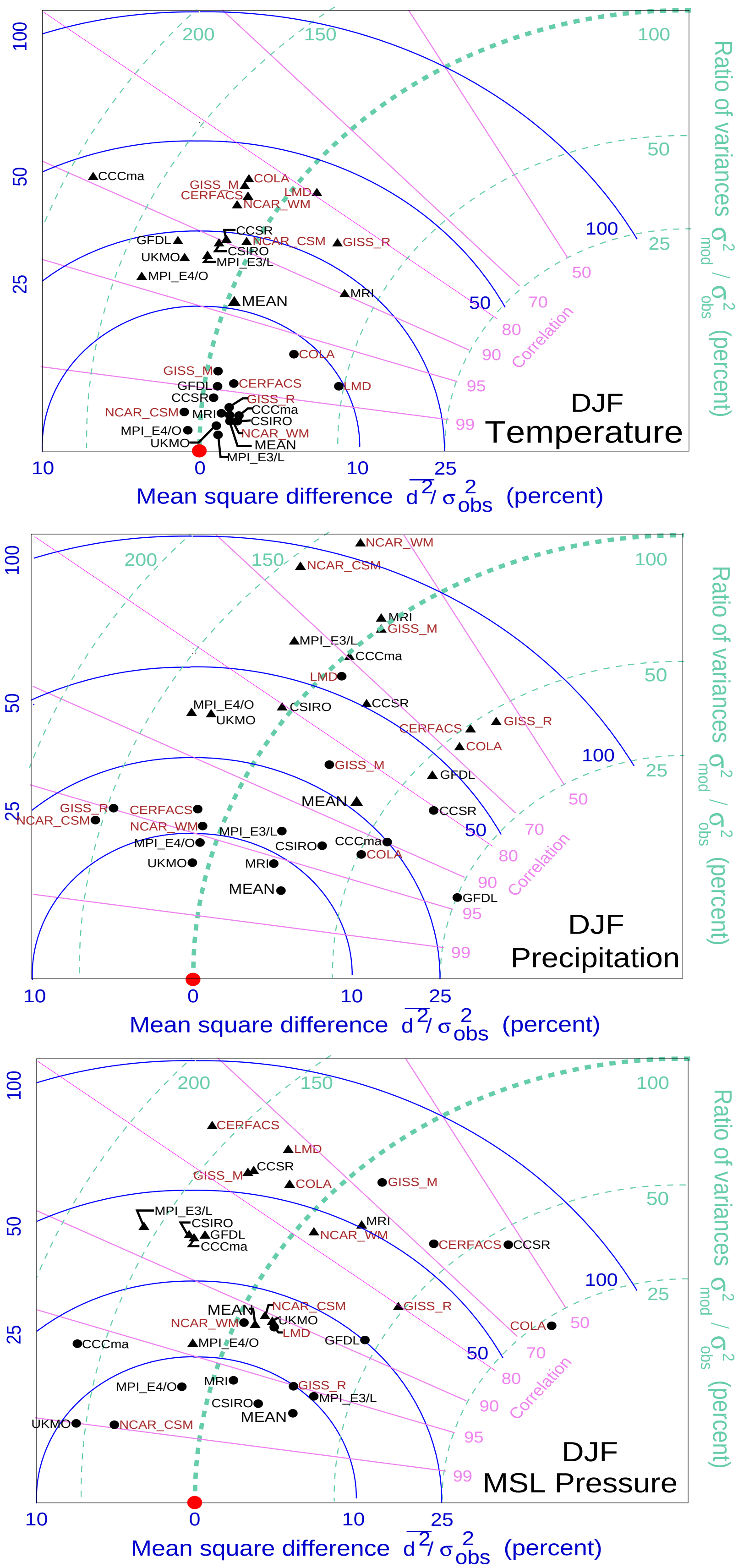

Figure 2. Second order statistics for individual models and the "mean model".
Dots are meridional and triangles the geographic structures

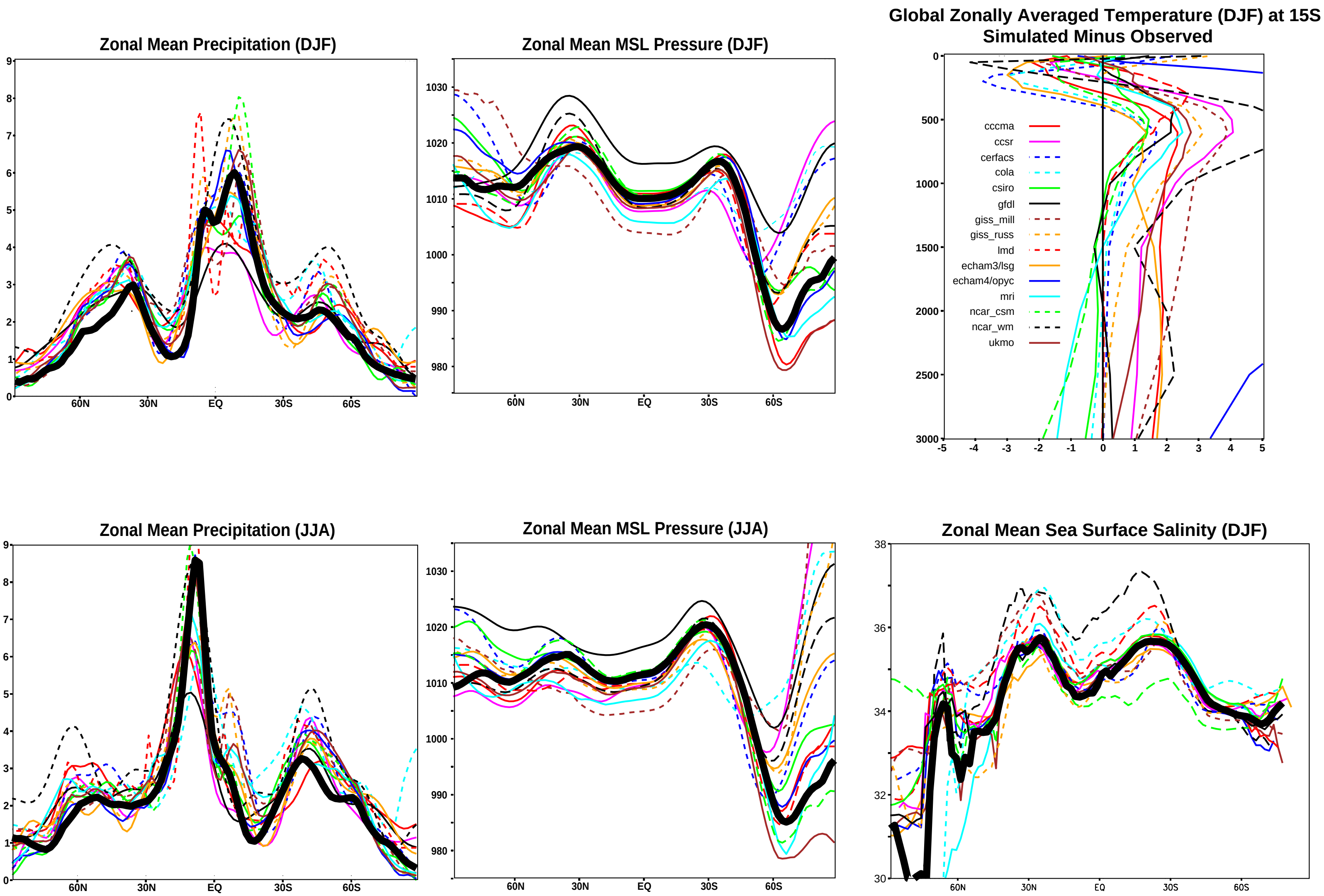

Figure 3. Zonal structures.