Authors comments on "Representation of the Tropical Stratospheric Zonal Wind in Global Atmospheric Reanalyses" by Y. Kawatani, K. Hamilton, K. Miyazaki, M. Fujiwara, J. Anstey

Corresponding author: Yoshio Kawatani (yoskawatani@jamstec.go.jp)

We are grateful to the four official referees for their helpful comments/suggestions and to Dr. Herzog for his contribution to the Open Discussion. We have revised the manuscript following their suggestions/comments. In this reply, we write the *reviewer's comments in blue italics*, while our responses are in regular fonts. At the end of this response are figures numbered R1-R6 which we prepared to help respond to the reviewers' comments, but do not propose to include in the revised manuscript.

Before presenting our detailed responses to each referee, let us address some overall issues that arose in some of the reviews.

Two of the four reviewers comment on the number of figures and panels in the manuscript. Reviewer #1 suggests some particular figures that could be removed or combined, while Reviewer #3 writes that "the number of figures is perhaps a little large" and suggests possibly reducing the total number of figure panels. We address the specifics in our responses to individual reviewers below where we explain that in many cases we prefer to keep the original figure. However, our efforts in this regard have resulted in a reduction of the 19 figures in the original version to 17 in our revision.

Some reviewers had suggestions for expanding the scope of our study in ways that would certainly be interesting, but which we feel are beyond the scope of our paper, which is strongly focused on what can be learned by intercomparing the monthly-mean zonal winds as represented in multiple state-of-the-art global reanalysis data sets. Notably Reviewer #4 suggests investigating the mean meridional circulation in the reanalyses noting the significance for chemical transport in the stratosphere. Abalos et al. (2015) have recently compared the stratospheric Brewer-Dobson circulation among MERRA, ERA-I and JRA55, while Miyazaki et al. (2015) did a somewhat similar study with 6 different reanalyses. Abalos et al. find quite large (~40%) differences among the reanalyses in the overall strength of the BD circulation and, specifically related to the QBO, they find that "There is a large spread among the estimates [of the fraction of variance in mean vertical motion explained by the QBO]". It is not clear that, even if we had the inclination to pursue this, we could add much to the Abalos et al. paper.

Extension of our study to include comparisons with special analyses that have been produced to test specific issues has been suggested. Reviewer 4 writes "I believe that it is beyond the scope of the paper to do assimilation experiments removing certain data sets from the analysis and in this way exploring the impact of particular data sets on the "message" of the reanalysis. But such studies have been done. And perhaps this paper could at least suggest ideas how to move forward in this direction" Along a similar line Reviewer 1 asks whether the JRA-55C reanalyses (which exclude all satellite data) could be included in our intercomparison. While we certainly agree that the comparison of analyses produced with subsets of the total input data can be illuminating, we feel this is beyond the scope of this paper and would to some extent duplicate the very recent study of Kobayashi et al. (2014) who directly compare the JRA-55C and JRA-55 reanalyses.

Reviewers #1 and #3 both raise concerns related to our treatment of the MERRA-2 reanalyses and wonder if it may be worthwhile to show more results directly comparing MERRA and MERRA-2. Since our original submission a very relevant new paper by Coy et al. (2016) has appeared in early online release for J. Climate. Coy et al. (2016) is a detailed look at the QBO as represented in MERRA-2 and includes a number comparisons to MERRA. They provide some answers to the question of how the enhancement in the gravity wave parameterization employed in the dynamical model for MERRA-2 affects the representation of the QBO. Coy et al find that "the increased equatorial gravity wave drag in MERRA-2 has reduced the zonal wind data analysis contribution compared to MERRA…" However Coy et al. also note the issues with fairly pronounced deficiencies in the early period covered by MERRA-2, specifically noting that before 1995 "MERRA-2 appears to overemphasize the annual cycle". This makes it problematic for us to include MERRA-2 in our calculation of the SD among all the other analyses over the long periods we considered. In our revision we have retained our basic approach which is to exclude MERRA-2 from our calculations except when showing the wind evolution during the rapid transition periods for individual reanalyses (Fig. 16 of revised version) composited just for the post-1998 period, where the comparison with MERRA is particularly relevant. Our conclusions on the likely role of the gravity wave drag in MERRA-2 seem to be consistent with the findings of Coy et al.

Reply to anonymous referee #1

The authors evaluate the representation of the stratospheric equatorial zonal winds from nine different reanalysis products against each other and against radiosonde observations, in particular the FUB wind record. They focus largely on the inter-reanalysis standard deviation (which is generally largest in the deep tropics) and comparison with radiosonde observations as a means of evaluating the reanalyses.

In nearly all cases they find that this standard deviation is anti-correlated with the spatial and temporal availability of radiosonde observations. In the mid-stratosphere (10 hPa), the standard deviation is dominated by the zonal mean component which is largest during transitions of the QBO phase where the reanalyses tend to lag the observations by 2 weeks to 2 months, particularly in the easterly-to-westerly transition. The eddy component correlates with the QBO phase, being larger during the westerly phase, and is apparently associated with the representation of extratropical stationary waves. Lower in the stratosphere and upper troposphere the eddy component becomes larger and the correlation with the QBO phase weakens. This structure appears to be associated with the stratospheric extension of the Walker circulation.

The analysis is a very useful contribution to the literature given the broad relevance of the QBO and the reliance of many studies on reanalysis products. The discussion is generally lucid and concise; my main criticism is that there are too many figures and it seems to me some of them could be removed (or combined) without impacting the central messages of the text. I would therefore recommend that the manuscript be accepted with minor revisions, either with a shortened figure list or stronger justifications in the text for those figures.

Thank you very much for your suggestion. As noted above we have tried to reduce/combine figures and have succeeded in reducing the total number of figures to 17 from the 19 in the original version of the manuscript.

The discussion could also be strengthened by including some comments regarding the implications of these results for (a) studies using the reanalyses to understand the QBO or its impacts and (b) reanalysis centres trying to improve the representation of the winds.

With regards to (a), the bias in timings of the phase transitions could be relevant for studies which *composite based on these dates, as could (possibly) the weak westerly wind maximum. The magnitude of the standard deviation in the horizontal winds throughout the tropics would also seem to be worth highlighting for trajectory studies given that there is some inclination to assume that all the inter-reanalysis differences are in the vertical velocities.*

These are useful points to emphasize and in the revision we have added these to the discussion in the explanation of Figs. 9 and 16 and in the summary session.

With regards to (b), one hypothesis that is raised is that having a forecast model with an internally generated QBO might reduce some of the biases. Given that MERRA 2 is in this category it seems that this hypothesis could be more explicitly tested; since it seems there are still significant biases, it would seem this is not sufficient to guarantee improved representation.

As we noted above Coy et al. (2016) have very recently examined this issue in some detail. They found that MERRA-2 has reduced the zonal wind analysis increments compared to MERRA, so that the QBO mean meridional circulation can be expected to be more physically forced and more physically consistent. In our revision we include these findings of Coy et al. in the discussion section.

Could the errors (particularly in the mid-stratosphere) be associated with the slow advective propagation of information from the Singapore winds during phase transitions?

This would give you larger errors during transitions. Is it more likely the reanalysis forecast models are systematically biasing the winds relative to radiosonde observations during transition periods?

These are two good suggestions that may help explain the lag in the reanalysis winds during the QBO transitions. We had already included the second possibility in our original manuscript, i.e. we had speculated that the delay could be understood as a consequence of the dynamical model bias (see discussion starting at page 9, line 21, in the original manuscript). In our revision we include also the possibility that during periods of weak zonal wind there will be delays in the zonal advective propagation of information introduced into the analysis system from observations at individual stations (notably Singapore). We are not able to definitely assess these two possible mechanisms, but the fact that the lag in the reanalysis winds is more pronounced in the more rapid easterly-to-westerly wind transition may favor the importance of the model bias over the slow advective propagation mechanism.

Perhaps relevant to both (a) and (b), if reanalyses (or free running GCMs) are going to use existing winds to nudge towards a QBO, which reanalysis would the authors recommend (or perhaps another way to ask the question, are there any that should be cautioned against?)

Our results support two strong conclusions in this regard: that the NCEP reanalyses are notably deficient in their representation of the QBO, and that systematic anomalies in the annual cycle contaminate the first part of the MERRA-2 record (a point now confirmed by Coy et al., 2016). Beyond that we have not tried to identify overall “best” analyses and have no results that strongly favor or disfavor application of individual data sets. However if the metric is how well the monthly mean zonal winds in the reanalyses compare with high quality direct station balloon observations, then ERA-I overall has a slight edge over the other data sets, as shown in Figs. 2, 9, 18 in the original manuscript (Figs. 2, 9, 16 in the revised manuscript). We add a comment “The RMS differences from FUB values in 1979–2012 are smallest in ERA-I, while those in NCEP-1 and NCEP-2 are much larger than those in the other reanalyses” in the discussion of Fig. 2f.

Specific Comments:
Figure 1 and 2 contain much the same information, but Figure 2 is more useful; the former could be omitted without losing any conclusions. Also the latter could be improved if the labels and titles of each panel in 2 were removed (say, using a single labeled time axis and annotations within the panel) so that the lines were more easily seen. Similarly Fig 6 contains the information in Fig. 5 and adds to it; Fig 5 can be omitted.

We considered the reviewer's suggestions here and also tried to reconsider the value of each of the figures in the original version. We decided to remove the original Fig. 4 (showing SD among reanalyses data sets as a function of height) as perhaps the least interesting to readers. The reviewer is correct that Figs. 1 and 2 present the same information. However, we decided to retain Fig.1 (simple height-time sections of zonal mean equatorial zonal-mean zonal wind) as the most basic presentation of the data sets considered. In principle, anyone could trivially duplicate this by simply grabbing publically available data sets and making contour plots, but we feel many readers will appreciate the considerable effort we put in to actually produce a single legible figure summarizing these data conveniently. The reviewer is also correct that the information in the shading of Fig. 5 is reproduced in the contours of Fig. 6. We arranged this so that the reader's attention could be first drawn to the overall structure of the SD in Fig. 5 without the complication of the station locations. We feel this allows for a logical exposition and sets up the reader to appreciate the more involved comparison in Fig. 6.

Fig 8: The time-dependence of the standard deviation would be much clearer if it was plotted on a different scale than the winds themselves. It also might be more informative to plot the standard deviations from each pressure level on the same axis so their temporal relationship can be more easily seen.

This is an excellent point. We changed the figure in accord with this suggestion and it is definitely clearer now.

Fig. 9 and 18 could easily be combined.

Indeed the panels from these two figures could be combined. However, given the discussion surrounding the original Fig. 18 (Fig. 16 of the revision) comes quite a bit later in the paper and concerns only the post-1998 data, we feel it is easier for the reader if we retain the figures with separate numbers.

Figs. 11 and 12 could also be combined and 11 (a-b) omitted.

We combined the original Figs. 11 and 12, but felt that 11a,b was useful for the reader and retained these panels as the a,b of the new figure (Fig. 11 in the revision).

p5 l21-25 It looks from Fig. 2d and e there are still significant anomalies in MERRA2 at 50 hPa and 70 hPa. Are these really still likely to be associated with an overactive SAO? This would seem to contradict the claim in l 26-27, though it is difficult to distinguish the MERRA and MERRA 2 curves.

As noted earlier Coy et al. (2016) addressed this and indicated that the deficiencies in the early period covered by MERRA-2 appear to be related to representation of the annual cycle. At 50hPa, the MERRA-2 winds are not so anomalous, however. It is the NCEP reanalyses that have larger apparent anomalies at 50 hPa.

p5 l29-30 What exactly was done with the tropical winds in NCEP-CFSR? Were they nudged towards ERA40 winds over this period? Is there a reference for this or will it be mentioned somewhere in the SRIP report?

Saha et al. 2010 mentioned in their section "QBO PROBLEM IN THE GSI" that "In order that the streams could proceed with a reasonable QBO signature, it was decided that the ERA-40 stratospheric wind profiles should be used as bogus observations for the period from 1 July 1981 to 31 December 1998". We add this reference here.

p6 l1-2 The justification for omitting MERRA 2 from much of the rest of the analysis is unclear to me, particularly since the authors return to it in Figs. 18 and 19. Why not just discuss it with the rest of the reanalyses? It also seems that NCEP-CFSR could be included for simplicity of method, though the case for omitting it is stronger. From Fig. 18 it looks like it does a good, if not better job of the transitions than other reanalyses; are the errors stronger in the middle of the QBO phase?

As noted earlier, we identified a peculiar anomaly in the representation of the annual cycle in the first part of the MERRA-2 record, a point now confirmed by Coy et al. (2016). This makes us reluctant to include MERRA-2 in our main results, namely calculations of SD among several analyses over long periods. However, given the interest in the accuracy of the reanalyses during the rapid transition phases, we did bring MERRA-2 into the discussion surrounding Fig 18 (Fig 16 in the revision), because the fact that the MERRA-2 dynamical model produces a QBO seems particularly relevant here. To avoid the problems with MERRA-2, of course, we limited the period considered in this one case to just post 1997. If we add MERRA-2 results to the original Fig. 9 (and adjust for the fact that MERRA-2 only starts in 1980) then we get the Fig. R6 below, where the MERRA-2 results really look peculiar.

With regard to NCEP-CSFR, once again the fact that these analyses were constrained to very nearly agree with ERA-40 above 30 hPa makes their inclusion into our calculations of SD among several analyses rather problematic.

p6 l12-13 One of the main conclusions from Fig. 3 would seem to be that the SD is improving amongst more modern reanalyses products (at least up to MERRA - I'm guessing this would change if MERRA 2 were included here?)

Indeed adding MERRA-2 makes the overall results worse (higher SD). This can be verified in Fig. R1 below, where we show how Fig. 3 changes when we include MERRA-2 (we have had to restrict the period to 1980-2012, as MERRA-2 does not begin until 1980).

p8 l10-15 Fig 13 a, b suggests that some of the structure in the eddy component in the tropical lower stratosphere might be associated with an extension of the Walker circulation - this is an interesting possibility and is distinct from issues of data availability. Since the phase of the QBO has not been considered in Fig. 13 would it make some sense to move the discussion on p 11 l 3-14 here? Also, as a test of this hypothesis, is the 70 hPa standard deviation correlated with the strength of the upper tropospheric Walker circulation?

In our revised version we have moved the discussion of Fig. 13 (Fig. 7 in the revision) to the last part of Section 3.2, as suggested.

Fig. R2 below shows the longitude-height cross section of the temporal correlation for the period

1979-2001 between the absolute value of the zonal wind (i.e., the strength of the zonal wind) and the the standard deviation among reanalyses averaged over 10ºN-10ºS (i.e., temporal correlation for which the two time series are |[u]| and SD calculated following Eq. 1).

In the upper part of the Walker circulation, the relatively high positive correlation is seen in the eastern hemisphere, while the correlation is relatively low in the central Pacific. In the mean state, the eddy component in the tropical lower stratosphere might be associated with an extension of the Walker circulation, but 70hPa standard deviation is relatively small (Fig.13b in the original manuscript), and it seems not correlated with the strength of the upper tropospheric Walker circulation. The large SD from the upper troposphere to the stratosphere in the central Pacific could be simply related to the fewer in-situ observations available there. In the middle stratosphere, the correlation is negative, corresponding to large SD during the phase transition of the QBO. We add this figure in Fig.7c in the revised manuscript.

p8 l24-26: Is there any correlation between the QBO phase and the number of radiosonde observations available?

We calculated the correlation between the number of radiosonde observations (Fig. 17 in the original manuscript) and FUB zonal wind at each height, and confirmed that there is no correlation between the QBO phase and observational numbers at all heights of 10 to 70 hPa.

p9 l10: This underestimation of the maximum westerly winds is one of the clearest biases and should be brought out more clearly in the conclusions (and possibly the abstract as well).

This issue is a little subtle, as the statement we made at this point in the manuscript applies to the maximum winds in a composite stretching over a finite period (6 months) from the transition date. This is not quite the same as a simple statement that the QBO westerly extremes are underestimated in the reanalyses, so we are reluctant to call this point out in the Abstract and Conclusion.

p9 21-25: This hypothesis could be evaluated explicitly here if the MERRA 2 winds were included. If they are not a good test of this hypothesis for some reason this could be explained here.

When we include MERRA-2 in the original Fig. 9 (and adjust for the fact that MERRA-2 only starts in 1980) then we get the Fig. R6 below, where the MERRA-2 results really look peculiar. We return to this issue in the discussion around Figure 18a (Fig. 16a of the revised version) that includes the later part of the MERRA-2 data record. This shows that the MERRA-2 reanalyses display an easterly-to-westerly phase transition at Singapore that is even more rapid than in the direct balloon observations. These results may indicate that the gravity wave sources in MERRA-2 are now excessive. We have extended our discussion here in the revised manuscript in the discussion of Fig. 16.

p10 l3-16: It's not clear to me why the authors have chosen to focus on a single case here - surely more robust conclusions could be drawn by looking at the wavenumbers of the composited eddy component of the standard deviation? Indeed it might be interesting to see a zonal wave number spectrum of the standard deviation at several levels.

The point is that the single case we showed is quite typical of the quasi-stationary wave behavior throughout the record. Fig. R3 shows longitudinal variations of the *u'* (deviation of the zonal wind values from the zonal mean) at 10hPa over the equator in each January from 1990 to 2000. Please note that Fig.10g in the original manuscript showed zonal wind *u* in 1996 and 1999 in the same panel, but here we show *u'* at each year separately.

Figs. R4 and R5 show *u'* at 10hPa in each of five reanalyses in those Januaries when the equatorial zonal mean zonal wind is westerly (i.e., $[\bar{u}]_{eq}$> 0) in 1992, 1993, 1995 and 1997.

So we have checked other years and found the monthly mean eddy structures in the middle stratosphere associated with quasi-stationary planetary waves are qualitatively similar among years. We explain this more clearly in the revised version of the paper in the discussion of Fig. 10.

p12 l30: It would be interesting to test the importance of the satellite observations for the tropical winds by including JRA55c which only assimilates 'conventional' observations. This would seem to be a good way to strengthen many of the conclusions in this section, and is exactly the kind of question for which it is perfectly suited.

We agree that insights can be found through comparisons of full reanalyses with versions that have had some data inputs withheld, and indeed the JRA-55C data provide such a possible comparison. As noted earlier, however, we think that adding such comparisons is beyond the scope of the present paper which is strongly focused on what can be learned by intercomparing the monthly-mean zonal winds as represented in multiple state-of-the-art global reanalysis data sets. We note also that the recent study of Kobayashi et al. (2014) comparing the JRA-55C and JRA-55 reanalyses presents results related to the stratospheric QBO in summary section.

p23 l24-33: Given the close resemblance of the structures in Fig. 19 to other structures we've seen in many of the figures, I think these conclusions could be made without showing the figure explicitly.

We feel that showing a measure of how well the most "up-to-date" reanalyses perform in the most recent period will be of interest to many readers, and so in our revision we retained Fig.19 (Fig. 17 in the revised version).

Reply to referee #2, Prof. Marvin A. Geller

This is a very nicely written paper, but of course, I do have some suggestions for its improvement. I think this paper could benefit from a short discussion of the data assimilation process in the introduction. Such a discussion is not needed by those familiar with the data assimilation process, but many users of assimilation use the resulting products without realizing that they are amalgams of data, the underlying model, and the statistical methods utilized. For this paper, I think it important at the beginning to indicate that the underlying model has its own climatology, and data, where present, nudges the resulting products toward observations. Also, unobserved quantities are adjusted to be consistent with the data being inserted together with the model climatology. To me, this together with the fact that as the Coriolis parameter tends toward zero the mass field constraint on the winds become weaker and weaker

Thank you very much for your suggestions. The core members of S-RIP are now preparing the detailed introduction of the reanalysis data, but this paper will not be available before this study will be published. So, in our revised version we have included a short discussion of the data assimilation process in the introduction as follows.

"Data assimilation is the technique for combining different observational data sets with a model, by considering the characteristics of each measurement and taking into account errors in both the measurements and the model (e.g., Kalnay, 2003). Advanced data assimilation schemes like the 4D-Var technique use the information provided by various measurements, such as radiosonde and satellite-derived measurements, and propagates it, in time and space, from a limited number of observable variables to a wide range of meteorological variables to provide global fields that are dynamically consistent and in agreement with the observations. Meteorological reanalyses have been conducted at operational centers using various approaches, which ingest a variety of observations over the period of each reanalysis product. Differences in the forecast model, assimilated measurements, and data assimilation technique used for producing reanalysis datasets can lead to differences in their representation of the mean state, variability, and long-term trend of atmospheric fields"

Another global comment is that reference should be made to Randel et al. (2004). Quoting from the summary section of that paper, "QBO variations in temperature and zonal wind are underestimated to some degree in most analyses, as compared to Singapore radiosonde data. The best results are derived from the assimilated datasets (ERA-40, ERA-15, METO, and NCEP, in that order) and only ERA-40 has realistic zonal wind amplitudes above 30 hPa. The use of balance winds in the Tropics (derived from geopotential data alone) is problematic for the QBO." The authors may wish to state to what extent they are updating those conclusions.

We have added the reference to Randel et al. (2004). We agree the use of balance winds in the tropics is problematic for the QBO and this largely explains why the SD peaks so strongly on the equator (less constraint on the winds from satellite temperature retrievals at low latitudes). Our brief review of the history noted that ERA-40 is an improvement over ERA-15 and our own results showed deficiencies in the NCEP reanalyses. So our results seem consistent with the conclusions of Randel et al. (2004). Now, other newer reanalyses are available. MERRA-2 is unique because this model can simulate the QBO internally but has some other problems (see Coy et al., 2016, and also our discussion at several other points in this Authors' Reply).

The following are some more detailed comments.
1. Page 2, line 8: Perhaps the paper by Yoo and Son might appear in time to be cited. It shows that the QBO exerts greater influence on the MJO than does ENSO. The reference is as follows: Yoo, C., and S.-W. Son, 2016: Modulation of the boreal wintertime Madden-Julian Oscillation by the stratospheric Quasi-Biennial Oscillation, Geophysical Research Letters, accepted.

We include the reference of Yoo and Son (2016) in the introduction of our revised manuscript.

2. Page 2, line 14: The authors may wish to add a reference to Naujokat (1986) who states, ""The first three stations were used to produce a data set for the levels 70, 50, 40, 30, 20, 15, and 10 mb, which should be representative of the whole circumference at the equator since all investigations have shown that longitudinal differences in phase are small enough to be ignored." The implication here, not specifically said, is that QBO amplitude differences among the stations are more substantial, and likely cannot be ignored. This statement seems to be consistent with the conclusions in Hamilton et al. (2004). The authors might then go on to indicate whether they feel the reanalyses can capture such asymmetries. They might if the extropical planetery waves during QBO westerlies are well treated. Otherwise, I doubt they will.

Indeed our current results are in general agreement with those Hamilton et al. (2004). Specifically, like Hamilton et al. we find modest, but significant, systematic zonal asymmetries in the QBO. We agree with Dr. Geller that this has implications for the application of the FUB "single station" record as it splices data from three separate stations during different epochs. In the present paper all the data we employ is from the post-1979 period when the FUB data were entirely from Singapore. As to the question of whether the planetary waves are well treated in the reanalyses, our contribution in this paper is to intercompare the reanalyses. We find an overall basic agreement among reanalyses, but considerable variation in the exact structure and amplitude of the zonal asymmetries (see the analysis of the SD presented in the manuscript and also Figs. R4 and R5 below). Here we explain as follows: "The high quality of these balloon data, and the close proximity of the stations to the equator, has led FUB series to be widely used, despite being based on only a single station each month (and despite modest inhomogeneities that the changes of station location may introduce into the record; see Section 3 below)"

3. Page 3 line 10: The situation is rapidly changing in that many models now produce spontaneous MJOs (i.e., GISS, CAM, etc.). Perhaps it would be better to say few GCMs used for reanalysis produce a spontaneous QBO.

We believe Dr. Geller means "QBOs" here, not "MJOs". In our revised version we have been more explicit and said that "because most GCMs display fairly steady, weak prevailing zonal winds due to failure to reproduce a spontaneous QBO".

4. Page 7, line 18: I believe it also depends on the climatology of the underlying GCM.

In our revised version we have added "In addition, the SD may also reflect differences among the climatologies of the GCMs used in the analysis systems" here.

5. Page 10, lines 30-31: Do the authors have any idea why this might be so?

The main point seems to be that the quasi-stationary waves propagating in from the NH extratropics are quite weak below 30 hPa, a result that is consistent with the earlier findings of Hamilton et al. (2004).

6. Page 14, lines 27-30: To what extent do the authors think this might affect the FUB QBO data set, which is often taken to represent the zonally averaged QBO?

Following our comments above, indeed the systematic zonal asymmetry in the zonal wind QBO should introduce discontinuities in the FUB data set which is spliced together from observational records at Canton Island, Gan and Singapore. However, as discussed in Hamilton et al. (2004), these stations are confined zonally to 73E-172W and the contrasts in the QBO near the equator within this sector are fairly small (see Hamilton et al.).

Again, I want to emphasize that this is an excellent paper. The figures are excellent, and clearly indicate the points being made.

We appreciate your evaluation and encouragement.

Reply to anonymous referee #3

This paper assesses the tropical stratosphere variability in contemporary (and not so contemporary) climate reanalyses. Focus has been restricted to the effects of observational inhomogeneities going into the datasets, primarily from radiosondes, and how they constrain the observed features of the quasi-biennial oscillation and their potential role in the differences seen between datasets. It is found *that the inter-model disagreement coincides to data poor areas, especially in the lower stratosphere. The reanalyses show good agreement over Singapore and show a progressive improvement at more recent times. Consistent and sizeable biases remain in the timing of the phases of the QBO especially during the easterly-westerly transitions at 10hPa.*

This is a timely and well-written paper which will well-complement forthcoming science-focussed papers on tropical stratosphere variability. The number of figures is perhaps a little large, but they are generally of good quality. If the number of figure panels were to be reduced, the ability to see fine features in the data would be really improved. I recommend publication pending due consideration of the points outlined below.

As described in our response too Reviewer #1, we have tried to reconsider the value of each of our figures and decided to remove Fig. 4 and combine the panels in Figs. 11 and 12 (to produce Fig. 11 of the revised version).

Main points:
I am unable to understand why it is difficult to establish what observations have gone into the various reanalyses. I would imagine the information is likely to be conspicuously posted on the individual Reanalyses Centres' websites or have been collated by other groups participating in the SPARC S-RIP project. I would have thought the Reanalysis Centres would find it particularly informative where (inter-model) differences are potentially coming from. This information should be included (or pointed to) in the paper, in some convenient way.

We have contacted professionals in each reanalysis center and several members of S-RIP, and confirmed that the situation is as follows. Among the reanalysis data sets considered in this paper, only MERRA provides day-to-day observation information including detailed geolocations in a gridded form. Making such information publicly available in a user friendly manner was not feasible due to the volume and the complexity of observation data. Friendly formats and tools are currently being developed or considered at some reanalysis centers for their future reanalysis products. Furthermore, differences between different reanalyses are not just related to which observations were actually used (though this might have the largest impact), but might also depend on how a given data set was used, e.g., what data-quality-control and bias-correction procedures were actually applied or not applied. These procedures are different in the details for different reanalyses.

Currently, two manuscripts are being prepared. One is the S-RIP 2016 interim Report where there is a chapter on "Description of the Reanalysis Systems" written by Wright et al., and the other is an ACP overview paper on the special issue on "The SPARC Reanalysis Intercomparison Project (S-RIP)" by Fujiwara, Wright et al. Researchers from all the reanalysis centers are also coauthors of these manuscripts, having provided very comprehensive depiction of all the reanalyses including the information not given in the reanalysis reference papers. Please see these manuscripts once they become available.

There is an inconsistent use of MERRA-2 data within the paper. As stated by the authors, MERRA-2 represents the ONE dataset whose forward model reproduces a QBO. This reviewer for one, would be especially keen to see that dataset assessed more throughout the paper.

As we discussed earlier, the inclusion of MERRA-2 is problematic for our main calculations which relate to the differences among several reanalyses over long periods. Since our original submission a very relevant new paper by Coy et al. (2016) has appeared in early online release for *J. Climate*. Coy et al. (2016) is a detailed look at the QBO as represented in MERRA-2 and includes a number comparisons to MERRA. However Coy et al. specifically call out the issues with fairly pronounced deficiencies in the early period covered by MERRA-2, noting that before 1995 "MERRA-2 appears to overemphasize the annual cycle". This makes it problematic for us to include MERRA-2 in our calculation of the SD among all the other analyses over the long periods we considered. In our revision we have retained our basic approach which is to exclude MERRA-2 from our calculations except when showing the wind evolution during the rapid transition periods for individual reanalyses (Fig. 16 of revised version, composited just for the post-1998 period) where the comparison with MERRA is particularly relevant. Some measure of the difficulty in including the earlier MERRA-2 data is apparent in the Fig. R6 where we have included added MERRA-2 to the transition composites computed over a long period.

Is there a reason why 10hPa was chosen to assess the QBO phase timings? Outside of those times, there actually appears to be a better correspondence between MERRA-2 and the balloon record.

We chose 10 hPa as presenting the biggest challenge for the reanalyses as the numbers of radiosonde observations are fewest and the SD among reanalyses is largest. We add this sentence in the explanation of Fig.9.

Conclusions referring to the wavenumber structure of the reanalyses' tropical wind need to be tempered a little. Figure 10 identifies a period where the high latitude stratosphere was particularly active. Other figures indicating wavenumber structures in longitude over a longer period of time refer to reanalyses differences (i.e. SD)

We cannot include all years considered but show one particular year. We did same analysis in other years and confirmed the structures of quasi-stationary planetary waves are similar in spite of different amplitude. We have confirmed that the characteristics, namely (i) the eddy component near the equator appears to be dominated by zonal wavenumber 1 and 2 quasi-stationary planetary waves propagating from mid-latitudes during westerly phase of the QBO, and (ii) the eddy components are very small over the equator in the easterly phase of the QBO, are qualitatively similar in different years. The fact that eddy components are large during westerly phase of the QBO is shown in Fig. 9f for 1979-2001. In our revision we have added a sentence here: "these characteristics seen in westerly and easterly phase of the QBO are qualitatively similar among other years".

Finally, there are a lot of details going into the (numerous) figures. Some of these details were difficult to pick up even when blowing the figures up on screen. Can the authors make sure the figures are not lossy and try to improve the clarity between different models. The authors may also consider looking for a colour-blind-friendly contour scheme. Perhaps the authors should assess if they really need to include all 19 figures (>100 panels) in the main article.

Thank you for these comments. We are not sure exactly which figures the reviewer is concerned about here. The most problematic may be Fig. 7 (Fig. 6 in the revised manuscript) which was assembled in Illustrator from many individual panels. Even when this figure is blown up to 400% (so one panel spans the whole viewable space on the screen), the curves and lettering are still reasonably crisp. For this figure we feel that it is useful for the reader to see all the panels together on one page, so we are not inclined to try to address any possible issue of crispness by making this (or other figures) span multiple pages. We feel the legibility of the figures is not a serious problem. We had not considered the issue of whether our color schemes are best for color blind readers. Perhaps in the near future this issue could be taken up by the journal publishers or the research community at large, and suitable guidance issued for authors.

Other points: (L28, P4) "...well-known summary..." (?)

In the revision we have removed this characterization.

(L9, P5) Presumably the interpolation to the FUB/IGRA data is done from the native reanalysis model resolution and not from the common ERA-I resolution mentioned in the previous sentence?

For the results presented we did interpolate from the common ERA-I resolution. There is no significant difference if we interpolate from the native reanalysis model resolutions.

(L29, P5) Can the authors please find a suitable reference for the statement that NCEPCFSR uses ERA40 winds in the tropics at and above 30hPA from 1 July 1981 to 31 December 1998. That is extraordinary!

We refer to Saha et al. 2010 here, who stated in their section "QBO PROBLEM IN THE GSI" that "In order that the streams could proceed with a reasonable QBO signature, it was decided that the ERA-40 stratospheric wind profiles should be used as bogus observations for the period from 1 July 1981 to 31 December 1998". In our revision we have added this reference here.

The information that the bogus data is used "at and above 30hPa" is from a personal communication and we have confirmed this by analyzing the datasets themselves (i.e. we found that there is almost no difference between ERA-40 and NCEP-CFSR values at and above 30 hPa). However, as Saha et al. (2010) did not explicitly mention the vertical range of the bogusing, we have made an appropriate modification to the sentence in our revision.

(L10, P6) As the tropics are the focus for the paper, it would make sense to limit the latitudes in figure 3 to something like 20-30 degrees, for example. I do not think the extratropical SD differences show anything interesting anyway.

We believe that showing the region outside the tropics provides a better context for the reader to see how exceptional the equatorial region is. In our revision we follow the reviewer's suggestion to the extent of reducing the latitude range shown in Fig. 3 from 90S-90N to 60S-60N.

(L20, P6) Why have the authors chosen levels below 100hPa in figure 4? They should explain the reasoning behind this. Not much is written about this figure.

As noted earlier we have removed the original Fig. 4 from the revised manuscript.

(L25, P6) Perhaps refer here to 'Indonesia' or the 'maritime continent' rather than the 'warm pool'. Also, the central (through to western) Pacific is where the most conspicuous SD values reside. This clarification may be important in pointing to other sources of model disagreement: in particular modes of (ocean-)atmosphere variability (e.g. ElNino-Modoki).

We followed this suggestion and in the revision we use "maritime continent" and "central Pacific" to denote these regions. .

(L32, P7) Bogota data is on panel j not h. Also the authors should consider doing a F-test to compare the differences in variances with and without IGRA data.

Thank you for noting this typo. We have conducted an F-test to see whether the SD of monthly values computed over periods with IGRA data can be regarded as significantly different from the SD computed over periods without IGRA data. The results show that for several stations the differences between periods with and without IGRA data indeed can be regarded as significant. Specifically the SDs at Seychelles, Thiruvananthapuram, Ascension and Abidjan display homoscedasticity distributions, whereas the SDs at Nairobi, Menado, San Cristobal, Bogota, Manaus and Belem have heteroscedasticity distributions with and without IGRA data. In general, the stations with the largest (smallest) SD difference between periods with and without IGRA data have heteroscedasticity (homoscedasticity) distributions.

(L21, P9) It is a pity that MERRA-2 is not shown here, whilst perhaps attempting to minimise the strong SAO signal (at 10hPa) afflicting a couple of years early in the time record. As that model has a spontaneous QBO it would be very interesting to see the timing of the E/W and W/E transitions. It would clearly be less reliant on analysis increments and sufficient observations to constrain the QBO phase progression.

When we include MERRA-2 in the original Fig. 9 (and adjust for the fact that MERRA-2 only starts in 1980) then we get the Fig. R6 below, where the MERRA-2 results really look peculiar. We return to this issue in the discussion around Figure 18a (Fig. 16a of the revised version) that includes the later part of the MERRA-2 data record. This shows that the MERRA-2 reanalyses display an easterly-to-westerly phase transition at Singapore that is even more rapid than in the direct balloon observations. These results may indicate that the gravity wave sources in MERRA-2 are now excessive. We have extended our discussion here in the revised manuscript.

(L8, P10) It is evident that all 5 reanalyses shown do actually show a positive anomaly near the maritime continent, although there is evident a local maximum in 3-4 of the datasets (probably highlighted due to contouring). Might also highlight (and reference) the fact a wave-1 warming occurred during December 1998 (and a wave-2 in February 1999)

This is related with your comments above. Again, we have confirmed the result is not changed qualitatively when we analyze other years, in agreement with the earlier results of Hamilton et al. (2004).

(L17, P10) subtitle: "Dependence of the ? Difference..."

In our revised version we have changed this subtitle to “Difference depending on the QBO phase”.

(L29, P10) "The overall larger SD in the westerly QBO phase, as compared to the easterly QBO phase,..."

In our revised version we followed your suggested change.

(L7, P12) "...has reported stratospheric..."; reference, Pers. Comm.? The sentences following this need to be looked at. It is mentioned that a change in SD occurs around 1998, but then it is mentioned that there was a bias in the forward model of JRA-25. But presumably the bias in the forward model will not be responsible for changes around 1998 (forward model should not change during the reanalysis period – unlike operational analyses)

Fujiwara et al. (2015) explained this in detail as “The radiative scheme used in the JRA-25 forecast model has a known cold bias in the stratosphere, and the TOVS SSU/MSU measurements do not have a sufficient number of channels to correct the model's cold bias; after introducing the ATOVS AMSU-A measurements in 1998, such a cold bias disappeared in the JRA-25 data product". In our revised version we have added references to Fujiwara et al. (2015) and to the relevant work of Onogi et al. 2007.

(L25..., p14) The authors should relax the statement referring to wavenumbers 1 and 2 dominating the 'eddy' zonal wind. Figure 10, mainly shows the zonal anomaly of zonal wind for January 1999 (a month sandwiched between two SSWs), so may not be representative of conditions at other times (e.g. 1979-2001)

As noted earlier we have reason to regard this as describing typical circumstances. Again a number of individual cases are shown in the figures R3, R4 an R5 below.

Reply to anonymous referee #4

General
This paper is part of a series of papers reporting on the comparison of the representation of different aspects in major global atmospheric reanalysis datasets; here the focus is on the tropical stratospheric zonal wind and the QBO. The paper works out in detail the agreement and disagreement between the reanalyses. It also analyses the impact of different observations on the quality of the reanalysis. I recommend accepting the paper for publication in ACP, after some critique (see below) has been taken into account in the revised version.

Thank you for your positive evaluation.

The paper states that most free-running GCMs have problems simulating a realistic QBO. While I agree that the QBO is still a challenge for GCMs, progress has been made in recent years. Models that have addressed the major shortcoming of GCMs in this respect, namely the representation of atmospheric waves that contribute to driving the QBO, in particularly waves with short vertical wavelengths have shown success in allowing a quasi-biennial periodicity to emerge. Of course parametrised wave drag is still required to generate a realistic QBO. I suggest discussing these recent developments (Orr et al., 2010; Anstey et al., 2016) in a bit more detail.

In the revised manuscript we have added a short sentence to the Introduction explaining these recent developments.

Further, the QBO induces a secondary meridional circulation, i.e. QBO variations in meridional and vertical winds (Punge et al., 2009). As this point is both important for tropical transport and intimately related to the QBO, I suggest extending the analysis in the paper somewhat to cover this aspect. I think this could be an important contribution of this paper.

Abalos et al. (2015) have recently compared the stratospheric Brewer-Dobson circulation among MERRA, ERA-I and JRA55, while Miyazaki et al. (2015) did a somewhat similar study with 6 different reanalyses. Abalos et al. find quite large (~40%) differences among the reanalyses in the overall strength of the BD circulation and, specifically related to the QBO, they find that "There is a large spread among the estimates [of the fraction of variance in mean vertical motion explained by the QBO]". It is not clear that, even if we had the inclination to pursue this, we could add much to the Abalos et al. paper. In our revised manuscript we reference Abalos et al. and include a sentence on this issue in the summary section.

On many instances the paper points out observations of differences and aspects of the reanalyses that are interesting to note. For example the finding that quasistationary waves differ significantly among reanalyses. However, it is more important to make progress on finding the reasons for differences between reanalyses. Any suggestions in the paper how to make progress in this direction would be very helpful. I believe that it is beyond the scope of the paper to do assimilation experiments removing certain data sets from the analysis and in this way exploring the impact of particular data sets on the "message" of the reanalysis. But such studies have been done. And perhaps this paper could at least suggest ideas how to move forward in this direction in a discussion. For example what could be important and relevant assimilation experiments to perform?

While we certainly agree that the comparison of analyses produced with subsets of the total input data can be illuminating, we feel this is beyond the scope of our present paper which is strongly focused on what can be learned by intercomparing the monthly-mean zonal winds as represented in multiple state-of-the-art global reanalysis data sets. Differences between individual reanalyses are not just related to which observations were actually used, but might also depend on how a given data set was used, e.g., what data-quality-control and bias-correction procedures were actually applied or not applied. These procedures are different in the details for different reanalyses, and we cannot obtain the detailed information about this (see our response on this issue to referee #3 above).

In summary, I think the paper could be improved with respect to some aspects in the revision. I think it will be a valuable contribution to ACP.

Thank you for your positive evaluation.

Minor issues

- *p. 3, l 6: 'most such models'?*

In the revision we changed this to the explicit "most GCMs".

- *p. 4, l. 5: there are also more recent publications on this point*

In the revised version we have added a reference to Abalos et al. 2015 and Miyazaki et al. 2015.

- *p 7., l 17: So is this only an expectation?*

In the revision we removed "expected" here, which makes our meaning clearer.

- *p 7., l 32: off?*

We meant "off the equator".

- *p 9, l 4: if one writes $-$ in LATEX, then one obtains proper minus signs*

We actually are using MS-Word for our manuscript, and we will be careful to make sure this looks OK in the final version.

- *p 10, l 7: change 'represents' to 'shows'*

In our revision we followed this suggestion.

· *p 10, l. 26: continuous over which time period?*

We mean continuous over the 1979-2001 period under discussion here (e.g. Fig. 6b).

· *p 11, l. 12: I agree it is interesting, but what is the conclusion here?*

We may be seeing the effects of the low numbers of in situ observations in both the upper troposphere and stratosphere in the central Pacific region.

· *p 11, l. 20: I agree it is interesting, but what is the conclusion here?*

As we mentioned in conclusion section, our study has provided detailed results showing that the high accuracy and high resolution wind measurements by in situ radiosondes provide very important constraints in the reanalyses of circulation in the tropical stratosphere.

· *p 11, l. 23: why is there a drop?*

As we mention a few sentences later, this drop is reasonably attributed to improved satellite observations (AMSU-U).

· *p 12, l 5: How likely is the possible reason?*

We cannot be definitive, of course, but the number of radiosonde observations did not dramatically increase around 1998 as shown in Fig. 15 (Fig. 17 in the original manuscript), thus implicating the upgraded satellite observations as the likely main cause..

· *p 12, l 7: change 'that' to 'on'*

We have made this change in the revised version.

· *p 12, l 8, 9: I think you mean the radiation code of the forecast model here.*

Yes the reviewer is correct. In our revision we have modified this sentence.

· *p 12, l 29: Discuss how this could be tested.*

As you indicated, doing assimilation experiments removing certain data sets from the analysis and exploring the impact of particular data sets is one possible way forward. Comparison between JRA-55 and JRA-55C might be possible, although it is beyond the scope of the present study and remains in the future. We add this sentence in the summary and concluding remarks section.

· *p 13, l 14: the three most modern reanalyses*

In our revision we have made this modification.

· *p 14, l 4: why only 'nearly'?*

We have confirmed the zonal wind in NCEP-CFSR is very similar to that in ERA-40, but not perfectly identical. The ERA-40 analyses were used as bogus data above 30 hPa in the NCEP-CFSR reanalyses.

· *p 14, l 12: why?*

As shown in Fig 11, the SD has zonally more non-uniform structure in the westerly phase of the QBO at 10 hPa, but the mean SD becomes more zonally uniform (Fig. 4 or original Fig. 5). The penetration of the Walker circulation, which is not related with the QBO phase, does not appear to influence the circulation in the 10-30 hPa range (Fig. 13 in the original manuscript).

· *p 14, l 17, 18: It also means that the constraint is really necessary. So there are problems with the underlying model, correct?*

Yes, that is our interpretation as well.

· *p. 16, Durre et al: abbreviate first names Fig 2, bottom panel: lines are difficult to disentangle*

In the revision we have made this suggested modification.

Reply to short comment by Dr. A. Hertzog

In this detailed study, Kawatani et al. (2016, hereinafter referred to as K16) present an intercomparison of monthly-mean zonal winds in the tropics in recent reanalyses. While focusing on monthly means is fully justified when tropical large-scale circulations (like the Quasi-Biennial Oscillation or quasi-stationnary planetary waves) are addressed, I would like to emphazise that such a time average ignores a significant fraction of the wind variability in the tropical lower stratosphere. Indeed, propagating disturbances associated with planetary waves trapped in the equatorial wave guide (e.g., Kelvin and Rossby-gravity waves) are essentially discounted in K16, even though reanalyses in principle have sufficient horizontal and temporal resolution to resolve most of these waves. It might thus be appropriate that K16 briefly discuss the implications of such time averaging on their results. Two references are provided hereinunder to that purpose, and also to bring up some elements on reanalysis agreement vs accuracy.

1. K16 show that the agreement between reanalyses in monthly-mean zonal winds has continuously improved since 1979. The standard deviation (SD) among reanalyses reaches zonal-mean values of ~1 ms^{-1} at 70 hPa in the last decade they studied (2001-2011), and never exceeds 1.8 ms^{-1} locally (their Figures 14 and 15). Despite the difficulties of constraining the tropical-stratosphere dynamics in atmospheric models with observations (which are recalled by K16), these SDs are quite impressive, as they are less than the assumed uncertainty associated with radiosonde winds in most models during the assimilation process. Yet, such good agreement likely does not apply to instantaneous reanalyses (i.e., without monthly average): Baker et al. (2014) (their Figure 2) have for instance shown that, in 2010, the zonal-mean SD between ECMWF operational analyses and NCEP GFS in zonal winds at 300 hPa is typically 3 ms^{-1}, and can reach values over 5 ms^{-1} over the eastern Pacific and Indian Oceans, i.e. at least three times the values reported in K16: according to K16, the SD at 300 hPa should be less than in the lower stratosphere (their Figure 1 and 13).

2. Away from regions with assimilated observations, an agreement between reanalyses does not necessarily mean an equivalent agreement with observations, even in the most recent decade. For instance, Podglajen et al. (2014, hereinafter referred to as P14), who compared reanalyzed winds with independent in-situ observations performed along long-duration balloon flights in 2010, have reported occurrences where ECMWF operational analysis, ERA-interim and MERRA products all agree, while the balloon observations depart from them (see their Figure 4). These events, which are associated with equatorial waves, induce discrepancies between reanalyses and observations that can reach values as large as 10 ms^{-1} and last for weeks. They once again tend to occur over areas with (very) few radiosounding stations: the Eastern Pacific and Indian Ocean. In these areas, observational increments are very low (about 1 ms^{-1} or less, see Figure 11 in P14), and the model dynamics is essentially running freely in the lower stratosphere.

It may therefore be worthwhile to warn the readers that they should not over-interpret encouraging figures regarding the improved agreement between reanalyses reported in K16.

In the revised manuscript, we explicitly caution against over-interpreting our results which are based solely on monthly mean fields. In the summary section, we explain as follows:
"The difference among reanalyses using twice daily data should be much larger than our SD based on monthly mean data (cf. Baker et al. 2014). Padglajen et al. (2014) compare reanalyses winds with independent *in situ* observations performed along long-duration balloon flights. They report that ERA-I and MERRA represent similar disturbances associated with equatorial waves, but reanalyses depart from the balloon observations. As the present study focus on monthly mean field, our analyses ignore variability with shorter time scales".

I finally note that the lower agreement between reanalyses during QBO shear phases reported in K16 likely has a counterpart in the agreement between reanalyses and observations, as discussed in P14. Shear layers indeed tend to reduce the vertical wavelengths of waves that propagate in the shear direction while they increase the associated horizontal-wind disturbances. The reduced wavelength means that the model resolution may become insufficient to properly resolve the wave disturbances, even though the wave signal is present in the assimilated observations (see for instance Figure 9 in P14).

Yes we agree that the effect of mean wind variations on the wavelength of vertically-propagating waves can be a contributor to the dependence we found of SD on the QBO phase.

Six figures we prepared for this reply to reviewers

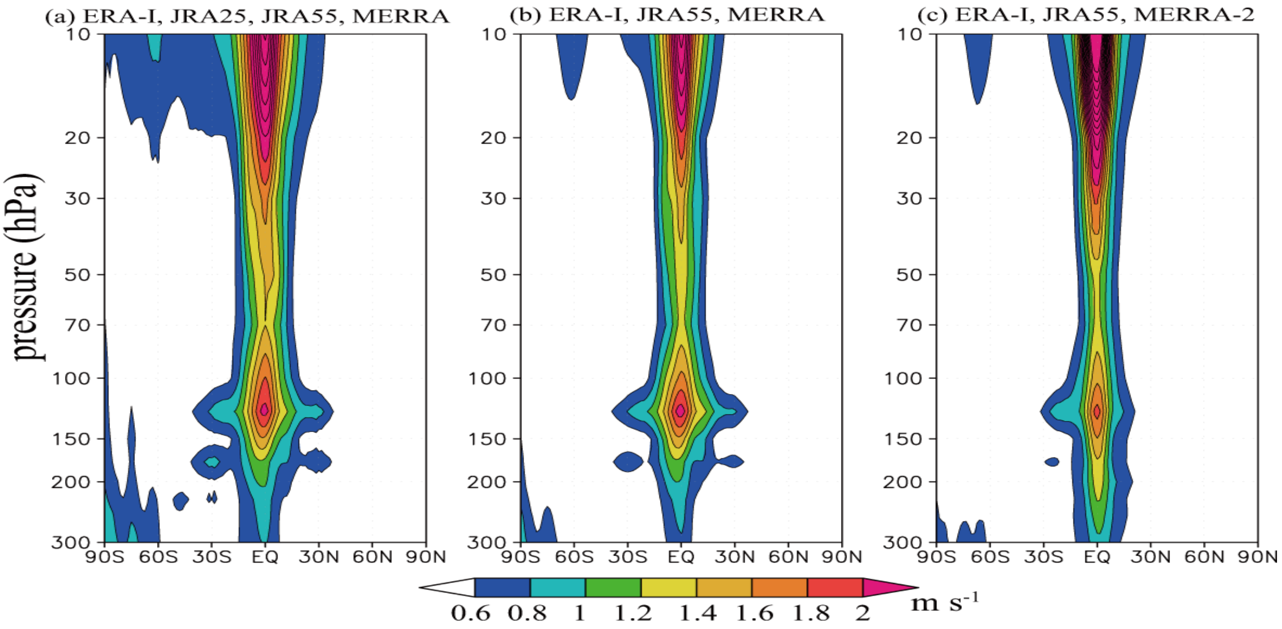


Fig.R1: Latitude-height cross-sections of the standard deviation among (a) four reanalyses of ERA-I, JRA-25, JRA-55, and MERRA, (b) three analyses of ERA-I, JRA-55, and MERRA and (c) three analyses of ERA-I, JRA-55, and MERRA-2 from 1980 to 2012.

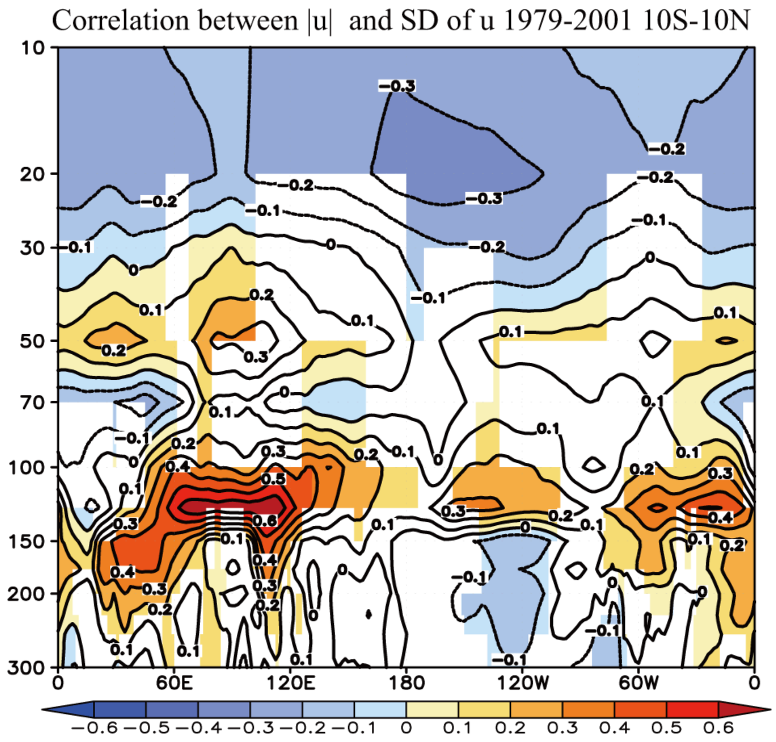


Fig. R2: Longitude-height cross section of the correlation between absolute value of zonal wind and the standard deviation among reanalyses in 10ºN-10ºS. Contour interval is 0.1 and correlations with 95% significance are shaded.

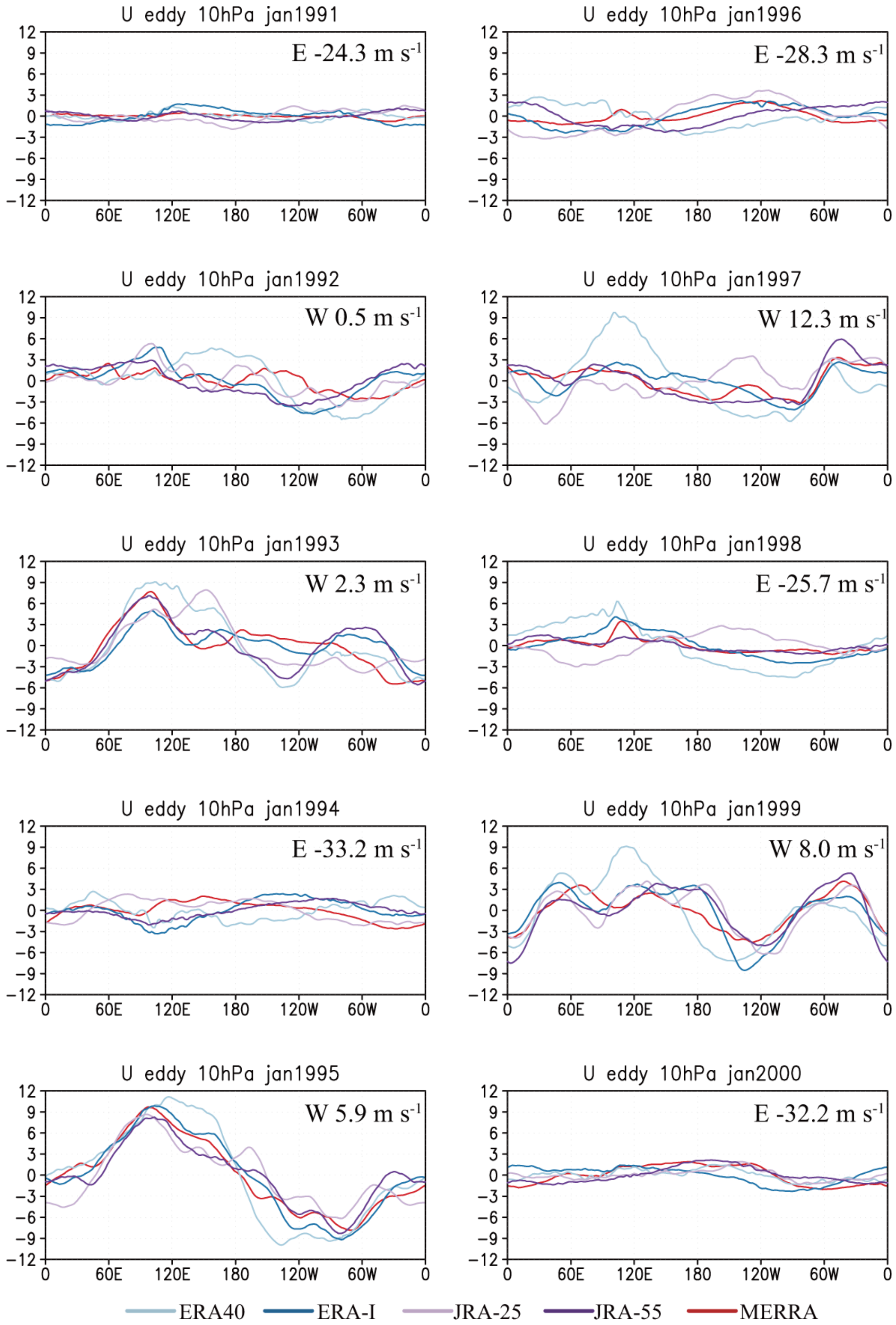


Fig. R3: Longitudinal variations of the *u'* (deviation of the zonal wind values from the zonal mean) at 10hPa over the equator in January from 1990 to 2000. The zonal wind values are shown on the top right corner. W and E mean westerly and easterly, respectively.

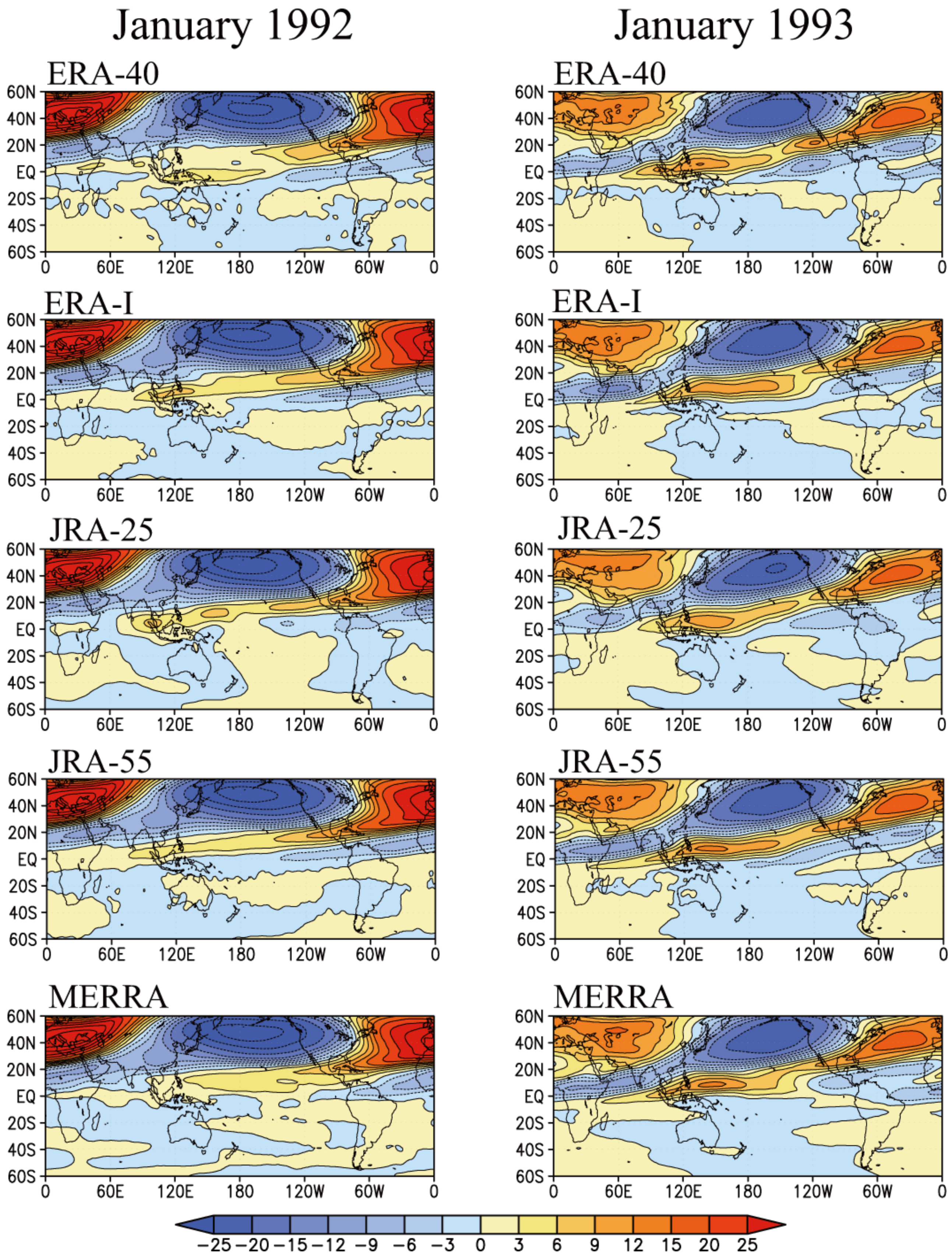


Fig. R4: Deviation of the zonal wind values from the zonal mean (*u'*) at 10hPa of five reanalyses in January (left) 1992 and (right) 1993 when the five reanalyses averaged equatorial zonal mean zonal wind is westerly over the equator.

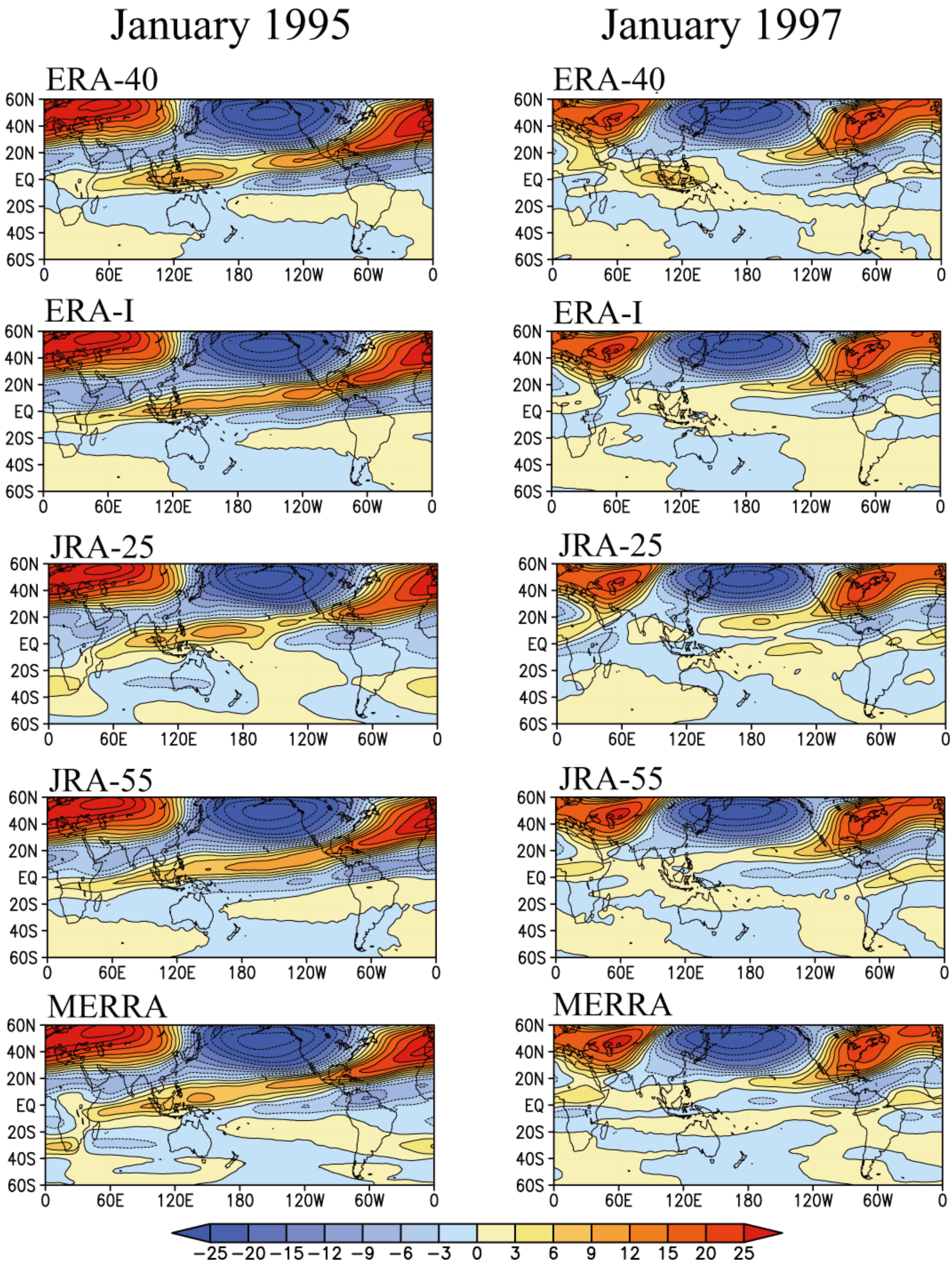


Fig. R5: Deviation of the zonal wind values from the zonal mean (*u'*) at 10hPa of five reanalyses in January (left) 1995 and (right) 1997 when the five reanalyses averaged equatorial zonal mean zonal wind is westerly over the equator.

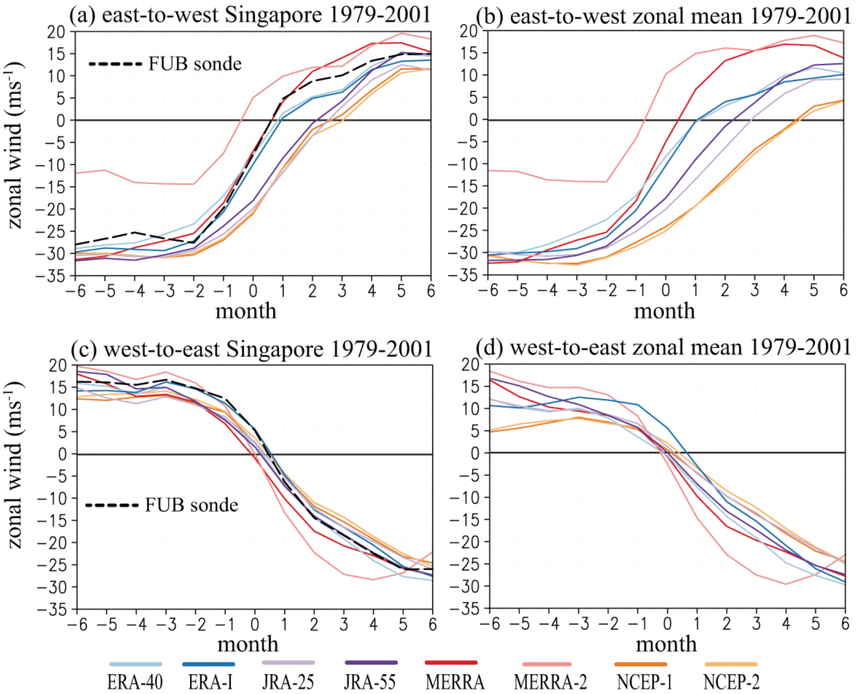


Fig.R6: The same as Figs.9a-d in the original manuscript but including MERRA-2 for 1979-2001. Due to overestimation of the SAO, 10hPa MERRA-2 zonal wind has large errors and it becomes difficult to discuss possible bias resulting from QBO phase transitions.

Representation of the Tropical Stratospheric Zonal Wind in Global Atmospheric Reanalyses

Y. Kawatani[1], K. Hamilton[2], K. Miyazaki[1], M. Fujiwara[3], J. Anstey[4]
[1]Japan Agency for Marine-Earth Science and Technology, Yokohama, 236-0001, Japan
[2]International Pacific Research Center, University of Hawaii, Honolulu, 96822, USA
[3]Faculty of Environmental Earth Science, Hokkaido University, Sapporo, 060-0810, Japan
[4]Canadian Centre for Climate Modelling and Analysis, University of Victoria, Victoria, V8W 2Y2, Canada

Correspondence to: Y. Kawatani (yoskawatani@jamstec.go.jp)

Abstract. This paper reports on a project to compare the representation of the monthly-mean zonal wind in the equatorial stratosphere among major global atmospheric reanalysis datasets. The degree of disagreement among the reanalyses is characterized by the standard deviation (SD) of the monthly-mean zonal wind and this depends on latitude, longitude, height and the phase of the quasi-biennial oscillation (QBO). At each height the SD displays a prominent equatorial maximum, indicating the particularly challenging nature of the reanalysis problem in the low-latitude stratosphere. At 50–70 hPa the geographical distributions of SD are closely related to the density of radiosonde observations. The largest SD values are over the ~~eastern Pacific~~central Pacific, where few *in situ* observations are available. At 10–20 hPa the spread among the reanalyses and differences with *in situ* observations both depend significantly on the QBO phase. Notably the easterly-to-westerly phase transitions in all the reanalyses except MERRA are delayed relative to those directly observed at Singapore. In addition, the timing of the easterly-to-westerly phase transitions displays considerable variability among the different reanalyses and this spread is much larger than for the timing of the westerly-to-easterly phase changes. The eddy component in the monthly mean zonal wind near the equator is dominated by zonal wavenumber 1 and 2 quasi-stationary planetary waves propagating from mid-latitudes in the westerly phase of the QBO. There generally is considerable disagreement among the reanalyses in the details of the quasi-stationary waves near the equator. At each level, there is a tendency for the agreement to be best near the longitude of Singapore, suggesting that the Singapore observations act as a strong constraint on all the reanalyses. Our measures of the quality of the reanalysis clearly show systematic improvement over the period considered (1979–2012). The SD among the reanalysis declines significantly over the record, although the geographical pattern of SD remains nearly constant.

1. Introduction

The dynamics governing the circulation in the tropical stratosphere have attracted much interest over the years (e.g. Wallace, 1973; Baldwin et al., 2001). As in other regions of the middle atmosphere, the day-to-day and higher-frequency variations of the flow in the tropical stratosphere are believed to be dominated by a spectrum of vertically propagating waves excited in the troposphere. What makes the low latitude stratosphere so remarkable is that the forcing of the zonal-mean flow by these waves leads to the very large-amplitude, low-frequency quasi-periodic cycle known as the quasi-biennial oscillation (QBO). In the tropical stratosphere, the QBO clearly dominates other aspects of interannual variability and even swamps the annual and semiannual variations in the zonal-mean circulation, at least up to ~3 hPa. Although rooted in the tropics, the QBO has global impacts. The QBO strongly influences interannual variations in circulation and composition throughout the stratosphere. The QBO also affects the circulation at the Earth's surface and is an important consideration in extended-range weather forecasts (e.g. Baldwin et al., 2001). Recently, Yoo and Son (2016) indicate the QBO exerts greater influence on the Madden-Julian Oscillation than does the El Niño-Southern Oscillation.

The state of the QBO up to the middle stratosphere can be characterized by the time series of monthly mean, near-equatorial zonal winds at levels between 10 and 70 hPa maintained by the Free University of Berlin (FUB, e.g. Naujokat, 1986) since 1953. The monthly values in the FUB series are based on operational balloon soundings, and the FUB record has been stitched together from such observations at Canton Island (2.8°S, 172°W from January 1953 to AugustMarch 1967), Gan (0.7°S, 73°E from September 1967 to December 1975), and Singapore (1.4°N, 104°E since 1976). The high quality of these balloon data, and the close proximity of the stations to the equator, has led FUB series to be widely used, despite being based on only a single station each month (and despite modest inhomogeneities that the changes of station location may introduce into the record; see Section 3 below).

Global atmospheric analyses that assimilate all available satellite remote sensing and *in situ* observations are another potential source of information about the QBO and other aspects of the circulation in the tropical stratosphere. Data assimilation is the technique for combining different observational data sets with a model, by considering the characteristics of each measurement and taking into account errors in both the measurements and the model (e.g., Kalnay, 2003). Advanced data assimilation schemes like the 4D-Var technique use the information provided by various measurements, such as radiosonde and satellite-derived measurements, and propagates it, in time and space, from a limited number of observable variables to a wide range of meteorological variables to provide global fields that are dynamically consistent and in agreement with the observations. Meteorological reanalyses have been conducted at operational centers using various approaches, which ingest a variety of observations over the period of each reanalysis product. Differences in the forecast model, assimilated measurements, and data assimilation technique used for producing reanalysis datasets can lead to differences in their representation of the mean state, variability, and long-term trend of atmospheric fields.

AHowever, a number of factors combine to make the global meteorological analysis process particularly challenging in the tropical middle atmosphere.

(i) One challenge is the relative paucity of *in situ* data, even if attention is restricted to levels at and below 10 hPa (i.e. the usual upper bound for most operational balloon soundings). There are large ocean areas in the tropics with no balloon observations. Even over land areas, the observations in the stratosphere at many stations are sparse. At stratospheric levels, the zonal wind measurements at many stations near the equator tend to have short overall records, or records with many months that have no observations (or not enough to compute a stable monthly mean). This leads to gaps in time series of monthly-mean winds (e.g. Hamilton, 1984; Kawatani and Hamilton, 2013).

(ii) Near the equator the Coriolis parameter is small; so observations of the temperature from satellite remote sensing do not constrain the wind field as strongly as at higher latitudes. Even if we assume the near-equatorial flow really is close to thermal wind-balance (Reed, 1962; Randel et al., 1999) the computed geostrophic wind shears are extremely sensitive to small errors in the observed temperatures.

(iii) The flow in the tropical stratosphere exhibits variations on very small vertical scales. This limits the usefulness of the relatively coarse-resolution satellite remote-sensing temperature retrievals. Even the monthly-mean zonal wind in this region displays thin layers where the wind can change by ~30 m s^{-1} over ~3 km. Satellite radiances used in global assimilations have an effective vertical resolution of several kilometers (e.g. Huesmann and Hitchman, 2003). Huesmann and Hitchman (2003) note that in such shear regions the assimilation scheme will have to reconcile the strong wind-shears measured directly by balloons with the somewhat weaker thermal winds consistent with the satellite derived temperature gradients (artificially damped due to the coarse vertical resolution).

(iv) Modern data assimilation approaches use general circulation model (GCM) simulations to obtain background (i.e. first guess) information and to determine data assimilation analysis increments. The tropical stratosphere is perhaps the only region of the atmosphere where most free-running GCMs have simulations with zero-order errors in the zonal-mean circulation (i.e. nothing even resembling a realistic QBO). This is because most GCMssuch models display fairly steady, weak prevailing zonal winds due to failure to reproduce a spontaneous QBO. Such models typically relax any QBO-like zonal winds in the initial condition towards the model climatology (e.g. Hamilton and Yuan, 1992; Saha et al., 2006; Boer and Hamilton, 2008) and so the persistent model bias could act against the data assimilation analysis increments and damp the QBO signals as introduced by data assimilation in the reanalysis fields. It should be noted that the situation is now slowly changing and several models have simulated a fairly realistic spontaneous the QBO in the equatorial stratosphere. Such models typically either employ fine vertical resolution as well as representation of momentum transports from small-scale non-stationary gravity waves either through high horizontal resolution (Kawatani et al. 2010 and reference therein) or by parameterization (Orr et al. 2010; Kawatani and Hamilton 2013; Anstey et al. 2016).

These special challenges in data assimilation for the tropical stratosphere have led to problems in actual global assimilation products, although there have been clear improvements over the last two decades in the ability of analysis systems to represent this region of the atmosphere. Trenberth (1992) documents a total misrepresentation of the QBO in the tropical stratosphere of ECMWF operational global analyses in the early 1980's, but finds that the situation improved considerably after changes in the assimilation scheme were introduced in May 1986.

In the 1990's some major meteorological centers started producing retrospective reanalysis products that employed multivariate statistical data assimilation methods such as optimal interpolation (OI) and three-dimensional variational (3D-VAR) scheme to combine all available data over some extended period. Such reanalyses have obvious advantages for research applications requiring the most homogeneous possible data set throughout an extended period. Two of the early major products were the ECMWF "ERA-15" reanalysis (covering 1979 to 1994) and the first NCAR/NCEP reanalysis (covering 1948–near present). The QBO in the tropical stratosphere in these products was examined by Pawson and Fiorino (1998, 1999) and Huesmann and Hitchman (2001, 2003) who found that the reanalyses displayed equatorial QBO variations that were clearly smaller in amplitude than in the real atmosphere. Randel et al. (2004) indicated that QBO variations in temperature and zonal wind were underestimated in most reanalyses available at that time, as compared to Singapore radiosonde data. They found that only ERA-40 had realistic zonal wind amplitudes above 30 hPa.

Some centers have now produced multiple reanalyses covering the same (or overlapping) periods; these more recent products are derived with updated data assimilation systems. Notable among these are the ECMWF ERA-40 (September 1958 to August 2002) and ERA-Interim (1979 to present). These products have improved representation of the QBO amplitude over that in ERA-15 (e.g. Pascoe et al., 2005; Baldwin and Gray, 2005). Baldwin and Gray (2005) compared the FUB zonal wind with ERA-40 data, and showed the zonal mean equatorial ERA-40 wind is quite close to that indicated in the FUB data.

Additional major reanalysis datasets have become available recently. The present paper reports on one of the studies contributing to the SPARC Reanalysis Intercomparison Project (S-RIP; Fujiwara et al., 2013; Errera et al., 2015), which are focused on evaluating reanalysis output for the stratosphere. We compare the representation of the circulation in the tropical stratosphere among several contemporary reanalysis products and validate the reanalyses against *in situ* observations. We have mostly limited our attention to just a single aspect of the circulation, namely the monthly-mean zonal wind. We limit our consideration to low latitudes (20°N–20°S) and a height range (10–70 hPa) in which the QBO is strong and where substantial balloon data are available for comparison. We have also restricted our attention to the period starting in 1979 when NOAA operational satellite radiance observations became available and were incorporated as an important data source in all the reanalyses. We appreciate that other aspects of the reanalyses in this region are also of interest, notably the ability of the reanalyses to represent accurately the large-scale transport circulation in the stratosphere (e.g. Coy and Swinbank, 1997; Abalos et al. 2015; Miyazaki et al. 2015), but these aspects will be left for future research.

Each reanalysis uses a different forecast model and assimilation scheme, and the types and numbers of assimilated observational data are also different among reanalyses. Furthermore, it is not feasible to determine exactly what observational data were actually assimilated at each data assimilation analysis step (e.g., what data quality control and bias correction procedures were actually applied). These complications make it somewhat difficult to attribute conclusively all the differences among the reanalyses products. However, it is interesting to investigate representations of key phenomena in the reanalyses. We hope such investigation will contribute to basic understanding and to improving future reanalysis products. We will show that, even with our somewhat narrow focus, some interesting conclusions concerning the representation of the tropical stratospheric circulation will emerge.

The outline of this paper is as follows. Section 2 briefly describes the reanalysis products that we evaluated and the station balloon data we employed. Section 3 investigates differences of the tropical zonal wind among reanalyses. Section 4 discusses evolution of the differences among reanalyses with time as different data sources become available. Section 5 summarizes the study and provides concluding remarks.

2. Reanalysis and radiosonde observation data

We analyzed monthly mean zonal wind and temperature in nine sets of global reanalyses data [NCEP-1 (Kalnay et al., 1996), NCEP-2 (Kanamitsu et al., 2002), NCEP-CFSR (Saha et al., 2010), ERA-40 (Uppala et al., 2005), ERA-I (Dee et al., 2011), JRA-25 (Onogi et al., 2007), JRA-55 (Kobayashi et al., 2015), MERRA (Rienecker et al., 2011) and MERRA-2 (Molod et al., 2015)]. Monthly mean data for 20°S–20°N and 10–70 hPa after 1979 are mainly analyzed, except for MERRA-2 which has data only after January 1980. Data before December 2012 are investigated, except for ERA-40 and NCEP-CFSR which are available until August 2002 and December 2010, respectively. With the exception of MERRA-2, none of the global dynamical models used in the assimilations would simulate the QBO when in free running mode. The dynamical model used in producing the MERRA-2 reanalyses is able to simulate a spontaneous QBO in the tropical stratosphere because it includes quite strong parameterized momentum fluxes from non-orographic gravity waves (Fig. 3 of Molod et al., 2015).

One key observational data set we employed for comparisons is the ~~well-known summary of~~ near-equatorial monthly mean values of operational balloon-borne radiosonde observations compiled by FUB (http://www.geo.fu-berlin.de/en/met/ag/strat/produkte/qbo/). The FUB data during 1979–2012 (i.e. period analyzed in this study) are based on Singapore observations.

We have also used monthly-mean values of the zonal wind at many other radiosonde stations provided in the Integrated Global Radiosonde Archive (IGRA: https://www.ncdc.noaa.gov/data-access/weather-balloon/integrated-global-radiosonde-archive, Durre et al., 2006). For 10°S–10°N IGRA includes balloon data from ~220 stations. When twice daily observations are available, the IGRA provides monthly means of observations at 00 and at 12 UTC separately. As in Kawatani and Hamilton (2013), where possible, the monthly values for 00 and 12 UTC are averaged, but the single values are used when only 00 or only 12 UTC data are available. The diurnal cycle of the wind in the lower stratosphere is expected to be small, and indeed we have found that generally the monthly mean zonal wind measured at 00 UTC data is nearly identical to that at 12 UTC.

In order to compare the data among reanalyses which have different spatial resolutions, all reanalysis variables are interpolated linearly to the ERA-I grid (i.e. 1.5° resolution in longitude and latitude, 37 vertical levels from 1000 to 1 hPa). When reanalysis data are compared with FUB/IGRA observational data, the reanalysis values are interpolated to the station locations.

3. Differences of the tropical zonal wind among reanalyses

3.1. Dependence of differences on longitude and height

Figure 1 shows the FUB zonal wind and the zonal wind in each reanalysis included in this study at Singapore (104°E, 1.4°N) from 1979 to 2012. All the reanalyses have a reasonably close resemblance to the FUB observations. Notably each reanalysis clearly captures the basic features of the QBO seen in the FUB data including the cycle-to-cycle variation in period and amplitude.

Figure 2 displays time variations of the zonal wind in each reanalysis at 10, 20, 30, 50, and 70 hPa over Singapore compared with the FUB observations. Also shown are the root mean square (RMS) differences between FUB and each reanalysis zonal wind averaged from 10 to 70 hPa. Reanalysis zonal winds are generally close to the FUB zonal wind. The RMS differences display somewhat noisy variations but an overall trend to smaller values over time is apparent. The zonal winds in NCEP-1 and NCEP-2 are generally underestimated, especially in the 1980's over 20–50 hPa. The most obvious anomaly in the reanalysis winds at 10 hPa is found in MERRA-2, which exhibits spurious semi-annual variations in the 1980's and in late 1993 particularly during easterly phase of the QBO (Figs. 1g and 2a). The downward propagation of westerly semiannual oscillation (SAO) phases is enhanced during these periods, which could~~would~~ be caused by overly strong~~overactive~~ gravity wave forcing (Fig. 3 of Molod et al., 2015). Coy et al. (2016) note that MERRA-2 appears to overemphasize the annual cycle before 1995.

The RMS differences from FUB values are smallest in ERA-I, while those in NCEP-1 and~~,~~ NCEP-2 ~~and MERRA-2~~ are much larger than those in the other ~~five~~ reanalyses. MERRA-2 represents large RMS differences in 1980's (Fig. 2f), ~~. In MERRA-2, the large RMS differences are~~ mainly due to the enhanced SAO at 10 hPa. On the other hand at 30–50 hPa, the MERRA-2 zonal winds show improved representation of the QBO compared to MERRA (Lawrence Coy et al.~~, private communication~~, 2016~~5~~). The NCEP-CFSR actually uses the ERA-40 stratospheric wind profile~~s~~ as bogus observations in the tropics ~~at and above 30 hPa~~ from 1 July 1981 to 31 December 1998 to obtain a reasonable QBO (Saha et al., 2010). We confirmed the differences between NCEP-CFSR and ERA-40 during this period ~~at these levels~~ are nearly zero. For these reasons, our study focuses mainly on the five reanalyses: ERA-40, ERA-I, JRA-25, JRA-55 and MERRA.

In order to quantify the spread among reanalyses, the standard deviations (SD) among the reanalyses are calculated as follows:

$$SD = \sqrt{\sum_{i}^{N} (u_i - [u])^2 / N} \qquad (1)$$

where i labels the individual datasets and there are N datasets included. The square brackets [ ] denote the mean over all N reanalyses. The SD is calculated for each month using monthly mean zonal wind. Because the ERA-40 data are only provided until August 2002, attention will sometimes be restricted to the SD during January 1979 to December 2001.

The latitude-height distributions of the zonal mean and time mean of the SD among five reanalyses (1979–2001) are displayed in Fig. 3a. Large SD is seen in the upper troposphere and the stratosphere at low-latitudes, while the SD in the extratropical regions is much smaller. Figures 3b and 3c depict the SD among four reanalyses (ERA-I, JRA-25, JRA-55, and MERRA) and among the three latest reanalyses (ERA-I, JRA-55, and MERRA) in 1979–2012. These are discussed in the next section. In any case, the distributions of the SD are quite similar.

In order to examine the SD in more detail later, the SD is divided into zonal mean components ($\bar{u}$; overbar denotes the zonal mean) and components of deviation from zonal mean (u': hereafter, referred to as eddy components). Substituting $u_i = \bar{u}_i + u'_i$ and $[u] = [\bar{u}] + [u']$ in Eq. (1) and taking the zonal mean yields:

$$\overline{(SD)^2} = \sum_i^N \overline{(\bar{u}_i - [\bar{u}])^2} / N + \sum_i^N \overline{(u'_i - [u'])^2} / N \qquad (2)$$

Here, the first term on the right side represents the zonal mean variance among reanalyses due to the zonal mean wind component and second term exhibits the zonal mean variance attributable to eddy components.

~~Figure 4 presents the vertical profiles of the zonal mean SD of all components, the SD due to zonal mean wind, and the SD from eddy components at 10°S–10°N among five reanalyses in 1979–2001. Around 150–70 hPa, the zonal mean SD due to eddy components is comparable to that due to zonal mean components, while zonal mean components dominate above 50 hPa.~~

~~Figure 5~~Figure 4 shows horizontal distributions of the SD at 10, 20, 30, 50, and 70 hPa. At 70–50 hPa, large SD is seen in the Indian Ocean, to the east of the ~~warm pool~~maritime continent, the Atlantic, and especially in the ~~eastern Pacific~~central Pacific. At 10–30 hPa, on the other hand, the SD becomes more zonally uniform with increasing height.

3.2. Dependence in the lower stratosphere

Satellite radiance observations are distributed much more uniformly and homogeneously than *in situ* balloon observations, so it might be expected that satellite data assimilation would not have a large effect on the spatial distribution of the SD among reanalyses. We may expect the SD geographical distributions to be related primarily to the availability of station observations. ~~Figure 6~~Figure 5a shows the locations of all radiosonde stations in the IGRA (magenta dots indicate the locations of the stations shown later in ~~Fig. 7~~Fig. 6). We calculated the fraction of months during 1979–2001 with valid monthly mean wind data at each level at each IGRA station. Figures 5~~6~~b–f display the fractional data coverage at each radiosonde station with at least some useful data from 10 to 70 hPa (the 100% coverage at Singapore in the FUB data is also indicated). The contours in these panels reproduce the values of the SD shown in ~~Fig. 5~~Fig. 4. Over 10°N–10°S, there are about 220 radiosonde stations (~~Fig. 6~~Fig. 5a), but only a fairly small fraction have significant data coverage at the stratospheric levels of interest here, particularly at 10 hPa.

At 70–50 hPa, distributions of the SD appear to be negatively correlated with the density of radiosonde observations. The SD shows the local minima in the zonal direction near the locations with high observational density, such as the vicinity of Sumatra, Borneo and the Malay Peninsula, and over much of South America. Large SD values are found over an extensive region in the eastern and central Pacific where few radiosonde observations are available.

Observational information is propagated with time and space through analysis steps and forecast steps in data assimilation. In analysis steps, the background error covariance matrix is used to determine the spatial structure and the magnitude of analysis increment. Advanced data assimilation techniques such as 4D-VAR (used in JRA-55 and ERA-I) are expected to provide more efficient constraints even for remote points, because of the use of flow-dependent analysis. The different data assimilation methods used in each reanalysis would~~ be expected to~~ contribute to the SD even if each reanalysis assimilated the same observational data. In addition, the SD may also reflect differences among the climatologies of the GCMs used in the analysis systems.

Next, we consider the time variations of the spread among reanalyses at individual places and its relation to the availability of local radiosonde data. Each panel in ~~Figure 7~~Fig. 6 shows the time variations of monthly mean 70 hPa zonal wind measured at a particular radiosonde station together with the values from each of the five reanalyses and the SD among the reanalyses. The blue/red colors of the SD curves in each panel denote times with/without radiosonde data at the station. In the upper right corner of each bottom panel, the total data coverage (black number) and the average value of the SD during periods with (blue number) and without (red number) radiosonde data at each station are shown.

At data rich stations with coverage ≥ 80% such as Thiruvananthapuram (~~Fig. 7~~Fig. 6c), Singapore (~~Fig. 7~~Fig. 6d), Chuuk (~~Fig. 7~~Fig. 6f), Majuro (~~Fig. 7~~Fig. 6g), and Ascension (~~Fig. 7~~Fig. 6m), reanalyses zonal winds generally agree well with the direct observations, and the SD among reanalyses is small. On the other hand, at Christmas Island (~~Fig. 7~~Fig. 6h) where no observational data are available from these 23 years, the SD is large throughout the record. There are no other stations with available data near Christmas Island (~~Fig. 6~~Fig. 5a), which may also contribute to the extended period of large SD there.

Menado (~~Fig. 7~~Fig. 6e) and San Cristobal (~~Fig. 7~~Fig. 6i) demonstrate that the reanalysis SD at the locations of near-equatorial stations (within 1.5° of the equator; Fig 5~~6~~a) can be appreciably reduced during times when radiosonde data from these stations are available. Bogota (~~Fig. 7~~Fig. 6j~~h~~) and Manaus (~~Fig. 7~~Fig. 6k) are good examples off the equator also showing smaller SD during months with balloon data than without these data.

At Nairobi (~~Fig. 7~~Fig. 6a), an unrealistic spike in the westerlies is apparent in the observed record for January 1981, but none of the reanalyses replicate this spike. This might be an example of data quality control (e.g. gross error check based on Observation-minus-Forecast departures statistics) removing observational data during the assimilation processes. In the evolution from late 1997 to early 1998, the direct observations at most stations indicate double westerly peaks, but the second peak, in early 1998, was not observed at Abidjan (~~Fig. 7~~Fig. 6n). ERA-I and ERA-40 data follow this pattern and indicate a single peak but JRA-25, JRA-55, and MERRA display double peaks. Consequently the SD has a local maximum during this period. These results suggest that the error check procedures or assimilation schemes employed in the different centers may respond differently to anomalous single station observations.

The correspondence between periods of availability of radiosonde data and SD values among the reanalyses is sometimes unclear, such as for at Seychelles (~~Fig. 7~~Fig. 6b), 1980–82 at Belem (~~Fig. 7~~Fig. 6l) and 1984–98 at Abidjan (~~Fig. 7~~Fig. 6n). However in general, the SD in ~~Fig. 7~~Fig. 6 is smaller during the months when radiosonde observations are available, and the SD tends to be larger when they are not available. These results suggest that the observations at these individual stations are having a significant influence constraining the reanalyses of the zonal wind in the tropical lower stratosphere.

To investigate what affects the longitudinal SD variation from the upper troposphere to the lower stratosphere, Fig 7 shows the longitude-height cross-section of zonal wind averaged over the five reanalyses and their SD (both averaged over the

1979-2001 period), as well as the temporal correlation for the period 1979-2001 between the absolute value of the zonal wind (i.e., the strength of the zonal wind, ‖[*u*]‖) and the SD. All panels show the 10°N–10°S average. In the upper troposphere Fig. 7a displays the familiar Walker circulation signal in prevailing zonal winds. Fig. 7a suggests the eddy components of the prevailing zonal wind at 70 hPa (and to some extent even at 50 hPa) can be regarded as part of the Walker circulation. Hamilton et al. (2004) showed in their general circulation model experiments, that below about 60 hPa, the deviation of the zonal wind from the zonal mean appears to be dominated by an extension of the tropospheric Walker circulation.

A zonally elongated region of large SD exists around 100–150 hPa, the upper part of the Walker circulation (Fig. 7b). This upper tropospheric SD is largest in the central Pacific, where few observational stations are located (Fig. 5a). It is interesting that the region of large SD around the central Pacific in the upper troposphere connects with that seen in the lower stratosphere.

The correlation between the absolute value of zonal wind and the SD among reanalyses shows relatively high positive correlation in the upper part of the Walker circulation in the eastern hemisphere, while the correlation is relatively low in the central Pacific (Fig. 7c). In the mean state, the eddy component in the tropical lower stratosphere might be associated with an extension of the Walker circulation, but 70hPa standard deviation is relatively small, and it seems not correlated with the strength of the upper tropospheric Walker circulation. The large SD from the upper troposphere to the stratosphere in the central Pacific could be simply related to the fewer *in situ* observations available there. In the middle stratosphere, the correlation is negative, indicating that the reanalyses disagree the most when the magnitude of the zonal wind is weakest. This corresponds to large SD during the phase transition of the QBO, as shown in the next section.

3.3. Differences in the middle stratosphere

In this subsection we discuss the SD at 10–30 hPa where generally fewer radiosonde observations are available. At 20–30 hPa, the SD becomes more zonally uniform compared with results at 50–70 hPa (~~Fig. 5~~Fig. 4). However, relatively higher-density observations around the Indonesian maritime continent, as well as over the North and South American and African continents do act to reduce the SD locally (~~Fig. 6~~Fig. 5c and 5~~6~~d). At 10 hPa, Singapore is the only station quite near the equator with high-density coverage, and several stations around 7–10°N in the western Pacific have data coverage of 40–80% (~~Fig. 6~~Fig. 5b).

Because the SD among reanalyses involves more zonally uniform structures in the middle stratosphere (~~Fig. 5~~Fig. 4), the zonally averaged ~~of total~~ SD ~~(all components of zonal mean, plus eddy components; black line in Fig. 4)~~ is investigated first. Figure 8 shows zonal mean equatorial zonal wind in each reanalysis plotted together with the zonally averaged SD among reanalysis at 10, 20, and 30 hPa. Large differences are seen when the QBO phase changes from easterly to westerly. In addition, the SD is generally larger during the westerly phase than during the easterly phase of the QBO. The dependence of SD on the QBO phase is most pronounced at 10 hPa.

To further investigate this dependence, a zonal wind composite based on the phase of the QBO was computed for the FUB observations and for each reanalysis over 1979–2001. For the easterly-to-westerly composite, the month '0' is defined as the final month with a mean easterly value in the 10 hPa FUB data and month '+1' is the first month with a mean westerly value. Composite values are then computed for ±6 months around these zero months (month 0 in August 1979, January 1982, March 1984…) for the easterly-to-westerly phases; see Fig. 2a. A similar westerly-to-easterly composite was also computed (with month 0 taken at June 1980, January 1983, June 1985…). We chose 10 hPa as presenting the biggest challenge for the reanalyses as the numbers of radiosonde observations are fewest and the SD among reanalyses is largest.

The QBO composites for the zonal wind at 10 hPa are shown in Figs. 9a–d. Results for the zonal wind at Singapore are shown in the left panels (a,c) while results for the zonal mean equatorial zonal wind are shown in the right panels (b,d). Through the easterly-to-westerly transition at Singapore (Fig. 9a), the composite zonal winds in most of the reanalyses have an easterly bias relative to the FUB observations particularly during months –1 to +3. This results in a delay in the time of the zero-crossing in the reanalyses relative to the FUB observations. The reanalyses differ considerably in terms of this bias, however. The delay is ~1.5–2 month in JRA-25, JRA-55 and ~0.5 month in ERA-40 and ERA-I. The MERRA zonal winds show good agreement with FUB data during the easterly-to-westerly transition. The zonal wind values in the reanalyses show better agreement with the FUB observations and closer agreement among themselves during the westerly-to-easterly phase transition (Fig. 9c).

In composites of the zonal mean zonal wind (Figs. 9b, d), during the six months following the transition date the maximum westerly in each reanalysis is slightly smaller than in the zonal wind over Singapore (~3–4 m s^{-1} for ERA-40 and ERA-I, ~2–3 m s^{-1} for JRA-25 and JRA-55, nearly the same for MERRA), while the maximum easterly is nearly same. In both phase transitions, the spread of reanalysis zonal mean zonal winds is larger than that over Singapore. It should be noted that the timing of both phase transitions in the zonal mean are generally the same as those seen just using results at Singapore (although there is a slight delay of ~0.5 months in the easterly-to-westerly transition of the zonal mean in the JRA-25 and JRA-55) .

Figures 9e and 9f show a QBO composite of the zonal mean equatorial SD among five reanalyses due to zonal mean and eddy components (square roots of the first and second terms on the right in Eq.(2), respectively) at 10 hPa. It is clear that the SD due to zonal mean components is large during the phase transition from easterly to westerly, while the SD does not change much during the 12 composite months during the westerly-to-easterly (Fig. 9e). The SD due to eddy components has larger values during the westerly than the easterly phase (Fig. 9f).

It is well known that the transition of the QBO from easterly-to-westerly is considerably faster than that from the westerly-to-easterly (e.g. Baldwin et al., 2001). In other words, the zonal wind tendency $|\partial u/\partial t|$ is larger during the easterly-to-westerly transition. As the QBO could not be simulated without data assimilation from each reanalysis model (except MERRA-2), one possible reason for the large SD in the easterly-to-westerly is weak forcing by resolved waves in the reanalysis model, leading to slow change of the zonal wind and resulting delay of the phase transition. The results of Fig. 9 suggest that the total observational constraints are insufficient to completely compensate for model bias, resulting in slow change of the zonal wind and delay of the phase transition. The another possibility is that during periods of weak zonal wind there will be delays in the zonal advective propagation of information introduced into the analysis system from observations at individual stations (notably Singapore). We are not able to definitely assess these two possible mechanisms, but the fact that the lag in the reanalysis winds is more pronounced in the more rapid easterly-to-westerly wind transition may favor the importance of the model bias over the slow advective propagation mechanism.

Each reanalysis model may not have sufficiently fine vertical resolution to represent completely the interaction between vertically propagating waves and the mean flow, which is thought to be crucial to the QBO dynamics (e.g. Baldwin et al., 2001), because the vertical wavelengths of waves become smaller as they approach critical levels. Podglajen et al. (2015) show vertical profiles of meridional wind by radiosonde observation at Singapore and in ECMWF operational analysis and indicate that ECMWF does not adequately represent wave disturbances due to insufficient vertical resolution (Fig. 9 of their study). MERRA has the finest ~~highest~~ vertical resolution among five reanalyses in the 20–30 km layer, which may allow a better representation of wave forcing in the strong vertical shears of the QBO. Kim and Chun (2015) investigate the momentum forcing of the QBO by resolved waves in reanalysis models, and show that MERRA has the larger net-resolved wave forcing than ERA-I, MERRA, and JRA-55 (Fig. 2 in their study). They also show that mean residual vertical velocity in the tropics, which generally acts in the opposite sense to wave forcing in driving the zonal wind acceleration (e.g. Kawatani et al., 2011), is much smaller in MERRA than in other reanalyses (Fig. 5a in their study). These are possible reasons why the MERRA shows faster transitions both in easterly-to-westerly and westerly-to-easterly transitions, compared with those in the other reanalyses.

Next we will discuss the representation of eddy components of the wind in the reanalysis datasets. Figures 10a–e show u', the deviation of the monthly mean zonal wind values from the zonal mean, at 10 hPa for each reanalysis in January 1999, during the westerly phase of the QBO. In all the reanalyses the eddy component near the equator appears to be dominated by zonal wavenumber 1 and 2 quasi-stationary planetary waves propagating from mid-latitudes, but the patterns differ somewhat among the reanalyses. For example, ERA-40 shows~~represents~~ a large, local, positive anomaly to the east of the Indonesian maritime continent, while JRA-55 does not. Each reanalysis shows a different shape for this positive anomaly. In addition, the negative anomalies around the Bay of Guinea and ~~eastern Pacific~~central Pacific are also different.

Figure 10f shows one example of the eddy components of the zonal wind in January 1996, during the easterly phase of the QBO in ERA-I. The eddy components are very small over the equator, which is expected as stationary planetary waves cannot propagate into the equatorial mean easterlies. Figure 10g depicts the longitudinal variations of each reanalysis zonal wind at 10 hPa over the equator in January 1996 and 1999. It is clear that the spread of the zonal wind among reanalyses is larger in the westerly phase, and longitudinal variation of the zonal wind is smaller in the easterly phase. We have checked other periods and found the monthly mean eddy structures in the middle stratosphere associated with quasi-stationary planetary waves are qualitatively similar among years (not shown). The large contribution to SD from eddy components during the westerly phase of the QBO (Fig. 9f) mainly results from different representation of mid-latitude planetary waves propagating into the equator among reanalyses.

3.4. Difference depending~~Dependence of the difference~~ on the QBO phase

In this subsection, we investigate the dependence of the SD among the reanalyses on the QBO phase. We have computed the SD separately for the westerly and easterly phases of the QBO defined by the sign of the five reanalyses averaged zonal mean zonal wind at the equator (i.e. defined as $[\bar{u}]_{eq} > 0$ for the westerly and $[\bar{u}]_{eq} < 0$ for the easterly phase; subscript *eq* indicates the equator). Figures 11a,b are maps of the SD among reanalyses at 10 hPa during the westerly and easterly phases of the QBO for 1979–2001. Figs. 11c-g plots the zonal variation of the SD averaged over 10°S–10°N at each height in the easterly and westerly QBO phases. At 10hPa, l~~L~~arge SD values in the westerly phase are seen in the Pacific, and from the Atlantic to Africa, where representations of quasi-stationary planetary waves differ significantly among the reanalyses (see Fig. 10). On the other hand, in the easterly phase, the SD is more zonally uniform. It is interesting that both in the westerly and easterly phases, the SD declines significantly around the longitudes near Singapore. At 10 hPa the Singapore station provides the only continuous record of *in situ* data near the equator (~~Fig. 6~~Fig. 5b), and the result in Fig. 11c suggest that observations from this one station act as a strong constraint on the reanalyses.

~~Figure 12 shows the same quantity as Fig. 11c but at 20, 30, 50, and 70 hPa.~~ The overall larger SD in the westerly QBO phases, as compared to the ~~than~~ easterly QBO phases, especially over the Pacific, is still clear at 20 hPa (Fig. 11d). However, this feature is not apparent at 30 hPa (Fig. 11e), suggesting that the effects of quasi-stationary planetary waves become much smaller at 30 hPa. At 50 hPa, the SD in the westerly and easterly phases is nearly identical (Fig. 11f). At 70 hPa the longitudinal variation of the SD in the westerly phase is similar to that in the easterly phase (Fig. 11g). This indicates that the QBO phase is not very important for SD zonal variations in the lower stratosphere.

~~To investigate what affects the longitudinal SD variation in the lower stratosphere, we show the longitude-height cross-section of averaged zonal wind from five reanalyses and the SD among reanalyses averaged from 1979 to 2001 and over 10°N–10°S in Fig. 13 (both easterly and westerly phases of the QBO are included). In the upper troposphere Fig. 13a displays the familiar Walker circulation signal in prevailing zonal winds. Fig. 13a suggests the eddy components of the prevailing zonal wind at 70 hPa (and to some extent even at 50 hPa) can be regarded as part of the Walker circulation. Hamilton et al. (2004) showed in their general circulation model experiments, that below about 60 hPa, the deviation of the zonal wind from the zonal mean appears to be dominated by an extension of the tropospheric Walker circulation, rather than meridonally propagating planetary waves.~~

~~A zonally elongated region of large SD exists around 100–150 hPa, the upper part of the Walker circulation (Fig. 13b). This upper tropospheric SD is largest in the eastern Pacific, where few observational stations are located (Fig. 6). It is interesting that the region of large SD around the eastern Pacific in the upper troposphere connects with that seen in the lower stratosphere.~~

4. Evolution of the differences among reanalyses with time

In this section, we discuss how the SD among reanalyses changes with time as the *in situ* and remote data sources available for the reanalyses evolve. We will show the results from the four reanalyses datasets (ERA-I, JRA-25, JRA-55, and MERRA) that exist over the entire 34 year period 1979 to 2012. Figures 12~~4~~a–c are maps of the SD averaged over 50–70 hPa for three 11-year intervals: 1979–1989, 1990–2000, 2001–2011. The SD among reanalyses reduces significantly with time, but it is interesting that geographical distributions of the SD remain quite similar over these three periods.

The left panels in Fig. 13~~5~~ show time variations of zonal mean SD among the four reanalyses at each height averaged over 10˚S-10˚N. At 70 hPa and 50 hPa (Figs. 13~~5~~j, m), the overall level of SD reduces gradually in the 1980s and 1990s, and a clear drop is found around 1998 at 50 hPa. At 10–30 hPa (Figs. 13~~5~~a,d,g), the SD also reduces with time, but a sudden drop is not apparent.

The middle panels in Fig. 13~~5~~ plot the same quantity as the left panels, but for the SD due to zonal mean and eddy components separately. At 70 hPa (Fig. 13~~5~~n), eddy components are comparable to zonal mean components for all periods, and both components reduce gradually with time until ~2000. The typical level of SD in both components remains fairly constant after 2000. At 50 hPa, the drop of SD in the zonal mean components around 1998 is pronounced; the eddy components also appear to become somewhat smaller after ~1998.

At 10–30 hPa, zonal mean contribution to SD always dominates, but gradually reduces with time (Figs. 13~~5~~b,e,h). However, throughout the record this quantity spikes generally near a particular phase of the QBO (i.e. during the easterly-to-westerly phase transition). The eddy contribution to SD also reduces slightly with time, but not dramatically.

One possible reason for reduction of the zonal wind SD over time would be upgrading of satellite radiance observations. From 1979 to 2006, the TOVS {TIROS (Television InfRrared Operational Satellite) Operational Vertical Sounder} Stratospheric Sounding Unit (SSU), and Microwave Sounding Unit (MSU), were available. After May 1998, data from the Advanced Microwave Sounding Unit (AMSU) became available. The advent of these new satellite measurements is a possible reason for the decline in the SD around 1998.

Figure 14~~16~~a shows the time-height section of zonal mean SD of temperature among the four reanalyses. A large SD is found around 30–50 hPa, and this SD reduces significantly after ~1998. The JRA team has reported ~~to us~~ on~~that~~ stratospheric temperature issues with JRA-25 and discussed the cause of the biases (Onogi et al. 2007). Fujiwara et al. (2015) explained that the radiative scheme used in the JRA-25 forecast model has a known cold bias in the stratosphere, and the TOVS SSU/MSU measurements do not have a sufficient number of channels to correct the model's cold bias; after introducing the ATOVS AMSU-A measurements in 1998, such a cold bias disappeared in the JRA-25 data product. ~~A cold bias existed in the forecast model of the radiation code used in producing JRA-25. The TOVS only had three SSU channels (centered at 15, 5, and 1.5 hPa) and one MSU channel (at 90 hPa), while AMSU-A has 15 channels including seven channels above 100 hPa. The influence of the TOVS measurements and radiosonde measurements did not completely correct the JRA-25 forecast~~

model bias at ~30 hPa. On the other hand, with JRA-55 using a new radiation scheme in the forecast model, the stratospheric temperature during the TOVS period has been much improved.

In light of the possible issues with the JRA-25 temperatures we show in Figure 16Figure 14b the temperature SD excluding JRA-25 (i.e. including only ERA-I, JRA-55, and MERRA). When JRA-25 is excluded, the very large SD values over 20–50 hPa before 1998 seen in Fig. 16Fig. 14a are greatly diminished. Even so the temperature SD among the three reanalyses (Fig. 16Fig. 14b) does reduce with time, with an especially large reduction over the full record at 100 and 10 hPa. The right panels of Fig. 15Fig. 13 show the same quantity as the left panels, but for zonal wind SD among just the ERA-I, JRA-55, and MERRA reanalyses. The SD among the three reanalyses is generally smaller than that among four reanalyses and the dramatic change in zonal wind SD at 50 hPa in 1998 become much weaker with the removal of JRA-25 (Fig. 15Fig. 13l), but the time evolution looks very similar.

Figure 17Figure 15 shows the evolution of the number of available monthly mean radiosonde observations over at 10°S–10°N and 10–70 hPa. The number of radiosonde observations available generally increased with time at all levels. For example at 70 hPa (50 hPa), the mean number of station-months of wind observations was ~33.7 (30.1) in 1979–1989, ~37.6 (35.1) in 1990–2000, and 44.5 (43.0) in 2001–2011 (i.e. periods corresponding to those in Fig. 14Fig. 12). So, relative to 1979–1989, the number of radiosonde observations at 70 hPa (50 hPa) increased by 11.6% (16.6%) in 1990–2000 and 32.0% (42.9%) in 2001–2011.

On the other hand, the zonal wind SD among the three reanalyses at 10°S–10°N (right panels in Fig. 15Fig. 13) and 70 hPa (50 hPa) reduces 20.8% (20.4%) in 1990–2000 and 33.7% (36.3%) in 2001–2011, compared with those in 1979–1989. The abrupt drop around 1998 is not clearly reflected in the number of radiosonde observations at 50 hPa, implying that the AMSU satellite observations could have significant impact on reanalysis of the zonal wind in the tropical stratosphere (see Figs. 135j, l).

At 10–30 hPa, the number of radiosonde observations increased significantly from 1979 to 2012, which should contribute to a reduction of the zonal wind SD with time at these altitudes, in addition to having more satellite data available. The same 10 hPa QBO composites as shown in Figs. 9a-d but during 1998–2012 are illustrated in Fig. 18Fig. 16. Here the zonal wind in ERA-40 is excluded (ERA-40 data exist only up to August 2002) but MERRA-2 is included (the unrealistic features noted earlier in MERRA-2 are not present after 1998). The general features are same as those in 1979–2001 such as an apparent delay in easterly-to-westerly transitions in JRA-25 and JRA-55 compared with the FUB observation and closer agreement among reanalysesan apparent advance in westerly-to-easterly transitions in most reanalyses, compared with the FUB observation. It should be noted here that in these reanalyses the 10hPa QBO in 1998–2012 are closer to the FUB observations both during easterly-to-westerly and westerly-to-easterly transitions than during 1979–2001 (Fig. 9). In addition, the spread of reanalysis zonal winds become much smaller in the later period, especially for the zonal mean wind. Figure 16 also shows that the MERRA-2 reanalyses display an easterly-to-westerly phase transition at Singapore that is even more rapid than in the direct balloon observations. These results may indicate that the gravity wave sources in MERRA-2 are now excessive (Coy et al. 2016).

Satellite radiance data will presumably affect the assimilated temperatures in the stratosphere, but will also have some influence on the wind (cf. Iida et al., 2014) mainly through the use of multivariate background error covariance matrix at analysis step, even though the coupling of temperature and wind may be somewhat weak as the Coriolis parameter is small near the equator. Thermal wind shears will be very sensitive to small errors in observed temperatures. After 2000 the amount of available satellite data increased greatly (e.g. Kobayashi et al., 2015), and the contributions of satellite radiance data to better representation of the tropical winds presumably also increased. However, it is hard to quantify the relative roles of global satellites and *in situ* radiosonde observations in reducing the zonal wind SD among reanalyses.

Finally, figures 17~~9~~a–d are the horizontal maps at 70–50 hPa SD among the three most modern ~~three~~ reanalyses (ERA-I, JRA-55, and MERRA) illustrated for 2001–2011, when the SD is smallest (~~Fig. 14~~Fig. 12). Here, we also calculated SD between ERA-I and JRA-55, ERA-I and MERRA, and JRA-55 and MERRA, separately. The SD becomes larger when JRA-55 data are included, and the SD between ERA-I and MERRA is smallest. However, in any case, the geographical patterns of the SD are always similar, independent of the reanalysis or period selected (see Figs. 4~~5~~, ~~14~~12, and ~~19~~17). These results indicate the importance of zonal wind observations by local radiosondes for better representation of the tropical zonal wind in the lower stratosphere in reanalysis datasets. Just for reference, the same figure, but between MERRA and MERRA-2, is shown in ~~Fig. 19~~Fig. 17e. The geographical pattern is also similar, even though the MERRA-2 assimilation was notable in using a dynamical model that simulated a spontaneous QBO.

5. Summary and concluding remarks

This paper reports on a project to compare the representation of the monthly-mean zonal wind in the equatorial stratosphere among major global atmospheric reanalyses data sets. We mainly confined our analysis to 20°N–20°S and 10–70 hPa and to data starting in 1979 (i.e. after the availability of operational satellite temperature soundings). We compared the zonal wind in the reanalyses with the Free University of Berlin (FUB) data set for equatorial monthly-mean winds, which, for this period, were based on balloon observations at Singapore. We also compared reanalysis values with balloon observations at many other low latitude stations included in the IGRA.

All nine reanalyses data sets we examined display monthly mean zonal wind values at Singapore that are reasonably close to the directly observed values at all levels between 70 and 10 hPa (Fig. 2). However, the NCEP-1 and NCEP-2 analyses have significantly larger differences with the Singapore observations than do the other reanalyses. Notably, NCEP-1 and NCEP-2 display considerably weaker (as much as 10 m s^{-1}) easterly QBO extremes than is apparent in the balloon observations, particularly during the 1980s. MERRA-2 also stands out as an outlier in displaying large differences from FUB at 10hPa in the 1980's and in November–December 1993. The NCEP-CFSR uses the ERA-40 winds in the tropical stratosphere ~~s at and above 30 hPa~~ from July 1981 to December 1998, and the differences between NCEP-CFSR and ERA-40 during this period ~~at these levels~~ are nearly zero. Further analysis and discussion here was largely restricted to the other five reanalyses (ERA-40, ERA-I, JRA-25, JRA-55 and MERRA).

We characterized the degree of disagreement among the reanalyses by the standard deviation (SD) of the monthly mean zonal wind values in the five data sets. This measure is a function of height, latitude, longitude, and time. At each height the SD displays a prominent equatorial maximum, indicating the particularly challenging nature of the reanalysis problem in the low-latitude stratosphere where *in situ* observations are relatively sparse and the Coriolis parameter is small. The SD in the tropical stratosphere also depends significantly on both longitude and height (~~Fig. 5~~Fig. 4). At 50–70 hPa, a large SD is seen for the Indian Ocean, to the east of the ~~warm pool~~maritime continent, the ~~eastern Pacific~~central Pacific, and the Atlantic, showing clear zonally non-uniform structures. At 10–30 hPa, on the other hand, the SD becomes more zonally uniform.

At 70–50 hPa, the distributions of the SD are closely related with those of *in situ* radiosonde observations (Figs. 5~~6~~e, f). The region of largest SD extends over wide zonal range in the ~~eastern Pacific~~central Pacific, where few observations are available. In the vicinity of individual radiosonde stations the SD at 70 hPa generally reduces in periods when observations are available relative to periods with no observations (~~Fig. 7~~Fig. 6).

At each level there is a tendency for agreement to be best near the longitude of Singapore (a result that holds in all QBO phases) suggesting that the high quality Singapore balloon observations act as a strong constraint on all the reanalyses. At 10 and 20 hPa, the zonal mean SD has a clear dependence on the phase of the QBO (Fig. 8). Specifically the SD generally jumps to 5–10 m s^{-1} for a few months around the transition from easterly to westerly. At 30 hPa, the difference is smaller (almost always less than 5 m s^{-1}) but it still shows the same QBO dependence. The main discrepancy with FUB observations at these heights is a tendency for the zonal wind accelerations to be delayed in the all reanalyses, except MERRA, around the easterly-to-westerly transition (Fig. 9).

At 10 hPa, all the reanalyses show that the eddy component (i.e. the deviation of the zonal wind values from the zonal mean) in the monthly mean zonal wind near the equator is dominated by zonal wavenumber 1 and 2 quasi-stationary planetary waves propagating from mid-latitudes (Fig. 10). Consistent with this, the eddy component of the zonal wind near the equator is small in months with zonal mean easterlies, and generally much larger in months with zonal mean westerlies. While the different reanalyses agree in broad terms about the structure of the eddy components in the middle stratosphere, there generally remains considerable disagreement among the reanalyses in the details of the quasi-stationary waves near the equator. At 50–70 hPa the SD distributions are not strongly affected by the phase of the QBO. The eddy features at these heights, rather than being planetary waves from mid-latitudes, might be related to stratospheric penetration of the Walker circulation (~~Fig. 13~~Fig. 7).

Our measures reflecting the reliability of the reanalyses show systematic improvement over the 1979–2012 period considered. Specifically, the agreement of the reanalysis results with the Singapore data improves significantly with time, with average deviations from the observation reduced by almost half from that in the early 1980s to the 2010s (Fig. 2f). At 70–50 hPa, the magnitude of the SD among the reanalyses declines significantly over the period of record (~~Fig. 15~~Fig. 13), although the geographical pattern of differences (as measured by the SD of the five reanalysis values) is nearly constant (~~Fig. 14~~Fig. 12).

For any individual reanalysis, the same atmospheric dynamical model and the same assimilation procedures are used throughout the whole period, thus any improvement seen in the quality with time must be related to changes in the observational data used. The number of available *in situ* radiosonde observations increase with time at all levels (Fig. 15~~7~~). In addition, satellite data streams vary throughout this period. The end of 1978 saw the introduction of the Stratospheric Sound Unit (SSU) which was flown on NOAA operational satellites for the next two decades (NOAA8 though NOAA14). We expect an important upgrade may have occurred in May 1998 when the temperature sounders on NOAA operational satellites were upgraded from the SSU to the Advanced Microwave Sounding Unit (AMSU). The satellite streams continued to show changes with the introduction of more AMSU instruments flying on other satellites. As well new sources of radiance data also appeared, such as the Atmospheric Infrared Sounder (AIRS) flown on the NASA Aqua satellite in 2002.

Doing assimilation experiments removing certain data sets from the analysis and exploring the impact of particular data sets are one possible way to investigate the data included. The effects of satellite datasets might be investigated by comparison between JRA-55 and JRA-55c (which excludes all satellite data as input). Kobayashi et al. (2014) compare the JRA-55 and JRA-55C reanalyses and indicates that the tropical zonal wind difference between the JRA-55 and JRA-55C is large in the upper stratosphere compared to that in the lower stratosphere, and the amplitude in JRA-55C is smaller than that of the JRA-55. More detail comparisons along these lines are reserved for future work.

The magnitude of the SD in the horizontal winds throughout the tropics would be worth highlighting for trajectory studies given that there is some inclination to assume that the key the inter-reanalysis differences for chemical transport are in the vertical velocities. Abalos et al. (2015) have recently compared the stratospheric Brewer-Dobson circulation among MERRA, ERA-I and JRA55, while Miyazaki et al. (2015) did a somewhat similar study with 6 different reanalyses. Abalos et al. find quite large (~40%) differences among the reanalyses in the overall strength of the BD circulation and, that there is a large spread among the estimates of the fraction of variance in mean vertical motion explained by the QBO. Coy et al. (2006) found that MERRA-2 which internally generates the QBO has reduced the zonal wind analysis increments compared to MERRA, so that the QBO mean meridional circulation can be expected to be more physically forced and more physically consistent

The difference among reanalyses using twice daily data should be much larger than our SD based on monthly mean data (cf. Baker et al. 2014). Padglajen et al. (2014) compare reanalyses winds with independent *in situ* observations performed along long-duration balloon flights. They report that ERA-I and MERRA represent similar disturbances associated with equatorial waves, but reanalyses depart from the balloon observations. As the present study focus on monthly mean field, our analyses ignore variability with shorter time scales.

Our study confirms that even with relatively few balloon stations, the high accuracy and high resolution wind measurements by *in situ* radiosondes have provided important constraints to reanalyses of circulation in the tropical stratosphere. Our analysis also shows that changes in both satellite and balloon data availability over the years have significantly affected the reanalysis of zonal wind data for the equatorial stratosphere. This conclusion has a negative aspect as it suggests that using reanalysis products (even restricting the data to after 1978) does not enable a researcher to avoid all significant artificial trends. On the positive side, it seems that the changes in available data have led to continually improving reanalysis representation of the tropical stratosphere.

Acknowledgement

We thank Prof. M. A. Geller, Dr. A. Hertzog and three anonymous reviewers for constructive comments on the original manuscript. We also express our gratitude to S-RIP members for providing useful information about the reanalyses data. This research was supported by Grant-in-Aid for Scientific Research B (26287117) and Joint international Research (15KK0178) from the Japan Society for the Promotion of Science, and by the Environment Research and Technology Development Fund (2-1503) of the Ministry of the Environment, Japan. This research was also supported by the Japan Agency for Marine-Earth Science and Technology (JAMSTEC) through its sponsorship of research at the International Pacific Research Center and by NOAA through grant No. NA11NMF4320128. The GFD-DENNOU Library and GrADS were used to draw the figures.

References

Abalos, M., Legras, B., Ploeger, F. and Randel, W. J.: Evaluating the advective Brewer-Dobson circulation in three reanalyses for the period 1979–2012, *J. Geophys. Res. Atmos.*,**120**, doi:10.1002/2015JD023182, 2015

Anstey, J. A., Scinocca, J. F., and Keller, M.: Simulating the QBO in an atmospheric general circulation model: sensitivity to resolved and parameterized forcing, *J. Atmos. Sci*., **73**, 1649–1665, 2016.

Baldwin, M.P. and Gray, L. J.: Tropical stratospheric zonal winds in ECMWF ERA-40 reanalysis, rocketsonde data, and rawinsonde data, *Geophys. Res. Lett*., **32**, L09806, 2005.

Baldwin, M. P., Gray, L. J., Dunkerton, T. J., Hamilton, K., Haynes, P. H., Randel, W. J., Holton, J. R., Alexander, M. J., Hirota, I., Horinouchi, T., Jones, D. B. A., Kinnersley, J. S., Marquardt, C.: Sato, K., and Takahashi, M.: The quasi-biennial oscillation, *Rev. of Geophys*., **39**, 179–229, 2001.

Baker,W. E., Atlas, R., Cardinali, C., Clement, A., Emmitt, G., Gentry, B. M., Hardesty, R. M., Källén, E., Kavaya, M. J., Langland, R., Ma, Z.,Masutani, M., McCarty, W., Pierce, R. B., Pu, Z., Riishojgaard, L. P., Ryan, J., Tucker, S., Weissmann, M., and Yoe, J. G.: LIDAR-MEASURED WIND PROFILES The Missing Link in the Global Observing System, *Bull. Am. Meteorol. Soc.*, **95**, 543–564, doi:10.1175/BAMS-D-12-00164.1, 2014.

Boer, G. and Hamilton, K.: QBO influence on extratropical predictive skill, *Clim. Dyn.*, **31**, 987-1000, doi10.1007/s00382-008-0379-5, 2008.

Coy, L. and Swinbank, R.: Characteristics of stratospheric winds and temperatures produced by data assimilation, *J. Geophys. Res*., **102**, 25763–25781, 1997.

Coy, L., Wargan, K., Molod, A., McCarty, W. and Pawson, S: Structure and Dynamics of the Quasi-Biennial Oscillation in MERRA-2. *J. Climate*, doi:10.1175/JCLI-D-15-0809.1, in press, 2016

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J-J., Park, B-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc*., **137**, 553–597, 2011.

Durre, I., Russell S. V.~~ose~~, and David B. W.~~uertz~~: Overview of the Integrated Global Radiosonde Archive. *J. Climate*, **19**, 53–68, 2006.

Errera, Q., Fujiwara, M., Long, C., and Jackson, D.: Report on the 10th SPARC data assimilation workshop and the 2014 SPARC Reanalysis Intercomparison Project (S-RIP) workshop. *SPARC Newsletter*, **44**, 31–37, 2015.

Fujiwara, M., and Jackson, D.: SPARC Reanalysis Intercomparison Project (S-RIP) Planning Meeting, 29 April–1 May 2013, Exeter, UK: *SPARC Newsletter*, **41**, 52–55, 2013.

Fujiwara, M., Hibino, T., Mehta, S. K., Gray, L., Mitchell, D., and Anstey, J.: Global temperature response to the major volcanic eruptions in multiple reanalysis data sets, *Atmos. Chem. Phys.*, **15**, 13507-13518, doi:10.5194/acp-15-13507-2015, 2015.

Hamilton, K.: Mean wind evolution through the quasi-biennial cycle of the tropical lower stratosphere. *J. Atmos. Sci*., **41**, 2113–2125, 1984.

Hamilton, K. and Yuan, L.: Experiments on tropical stratospheric mean wind variations in a spectral general circulation model. *J. Atmos. Sci*., **49**, 2464–2483. 1992.

Hamilton, K., Hertzog, A., Vial, F.: and Stenchikov, G., Longitudinal variation of the stratospheric quasi-biennial oscillation, *J. Atmos. Sci.*, **61**, 383–402, 2004.

Huesmann, A. S., and Hitchman, M. H.: The stratospheric quasi-biennial oscillation in the NCEP reanalyses: Climatological structures, *J. Geophys. Res*., **106**, 11859–11874, 2001.

Huesmann, A. S., and Hitchman, M. H.: The 1978 shift in the NCEP reanalysis stratospheric quasi-biennial oscillation, *Geophys. Res. Lett*., **30**, 1048, doi:10.1029/2002GL016323, 2003.

Iida, C., Hirooka, T., and Eguchi, N.: Circulation changes in the stratosphere and mesosphere during the stratospheric sudden warming event in January 2009, *J. Geophys. Res. Atmos.*, **119**, 7104–7115, 2014.

Kalnay, E.: Atmospheric Modeling, Data Assimilation and Predictability, Cambridge University Press, 341 pp., 2003.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo¸ K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-year reanalyses project, *Bull. Am. Meteorol. Soc*., **77**, 437–471, 1996.

Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S-K., Hnilo, J. J., Fiorino, M., and Potter, G. L.: NCEP–DOE AMIP-II Reanalysis (R-2), *Bull. Amer. Meteor. Soc.*, **83**, 1631–1643, 2002.

Kawatani, Y., Sato, K., Dunkerton, T. J., Watanabe, S., Miyahara, S. and Takahashi, M.: The roles of equatorial trapped waves and internal inertia-gravity waves in driving the quasi-biennial oscillation. Part I: Zonal mean wave forcing, *J. Atmos. Sci.*, **67**, 963–980, 2010.

Kawatani, Y., Hamilton, K., and Watanabe, S.: The quasi-biennial oscillation in a double CO_2 climate, *J. Atmos. Sci.*, **68**, 265–283, 2011.

Kawatani, Y. and Hamilton, K.: Weakened stratospheric quasibiennial oscillation driven by increased tropical mean upwelling, *Nature*, **497**, 478–481, 2013.

Kim, Y.-H. and Chun, H.-Y.: Momentum forcing of the quasi-biennial oscillation by equatorial waves in recent reanalyses, *Atmos. Chem. Phys*., **15**, 6577–6857, 2015.

Kobayashi, C., Endo, H., Ota, Y., Kobayashi, S., Onoda, H., Harada, Y., Onogi, K. and Kamahori, H.: Preliminary results of the JRA-55C, an atmospheric reanalysis assimilating conventional observations only. *Sci. Online Lett. on the Atmos*., **10**, 78-82. doi: 10.2151/sola.2014-016, 2014

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Hirokatsu, O., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K. and Takahashi, K.: The JRA-55 Reanalysis: general specifications and basic characteristics, *J. Meteor. Soc. Jpn*., **93**, 5–48, 2015.

Miyazaki, K., Iwasaki, T., Kawatani, Y., Kobayashi, C., Sugawara, S., and Hegglin, M.: Inter-comparison of stratospheric mean-meridional circulation and eddy mixing among six reanalysis datasets, *Atmos. Chem. Phys. Discuss*., **15**, 27749-27803, doi:10.5194/acpd-15-27749-2015, 2015.

Molod, A., L. Takacs, M. Suarez, and J. Bacmeister: Development of the GEOS-5 atmospheric general circulation model: Evolution from MERRA to MERRA2. *Geosci. Model Dev*., **8**, 1339–1356, 2015.

Naujokat, B.: An update of the observed quasi-biennial oscillation of the stratospheric zonal wind over the tropics, *J. Atmos. Sci*., **43**, 1873–1877, 1986.

Onogi, K., J. Tsutsui, J., Koide, H., Sakamoto, M., Kobayashi, S., Hatsushika, H., Matsumoto, T., Yamazaki, N., Kamahori, H., Takahashi, K., Kadokura, S., Wada, K., Kato, K., Oyama, R., Ose, T., Mannoji, N., Taira, R.: The JRA-25 reanalysis, *J. Meteor. Soc. Jpn*., **85**, 369–432, 2007.

Orr, A., Bechtold, P., Scinocca, J., Ern, M., and Janiskova, M.: Improved middle atmosphere climate and forecasts in the ECMWF model through a nonorographic gravity wave drag parameterization, *J. Climate*, **23**, 5905–5926, 2010.

Pascoe, C. L., Gray, L. J., Crooks, S. A., Juckes, M. N., and Baldwin, M. P.: The quasi-biennial oscillation: Analysis using ERA-40 data, *J. Geophys. Res*., **110**, D08105, 2005.

Pawson, S., and Fiorino, M.: A comparison of reanalyses in the tropical stratosphere, Part 2: the quasi-biennial oscillation. *Clim. Dyn*., **14**, 645–658, 1998.

Pawson, S., and Fiorino, M.: A comparison of reanalyses in the tropical stratosphere, Part 3: Inclusion of the pre-satellite data era, *Clim. Dyn*., **15**, 241–250, 1999.

Podglajen, A., Hertzog, A., Plougonven, R., and Žagar, N.: Assessment of the accuracy of (re)analyses in the equatorial lower stratosphere, *J. Geophys. Res*., **119**, 11,166–11,188, doi:10.1002/2014JD021849, 2014.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G-K., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C. R., Reichle, R., Robertson, F. R., Ruddick, A. G., Sienkiewicz, M., and Woollen, J.: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Climate*, **24**, 3624–3648, 2011.

Randel, W. J., F. Wu, F., Swinbank, R., Nash, J., and O'Neill, A.: Global QBO circulation derived from UKMO stratospheric analyses, *J. Atmos. Sci.*, **56**, 457–474, 1999.

Randel, W. J., Udelhofen, F., Fleming, E., Geller, M., Gelman, M., Hamilton, K., Karoly, D., Ortland, D., Pawson, S., Swinbank, R., Wu, F., Baldwin, M., Chanin, M., Keckhut, P., Labitzke, K., Remsberg, E., Simmons, A., and Wu, D.: The SPARC intercomparison of middle-atmosphere climatologies, *J. Climate*, **17**, 986–1003, 2004.

Reed, R. J.: Evidence of geostrophic motion in the equatorial stratosphere, *Q. J. R. Meteorol. Soc*., **88**, 324–327. 1962.

Saha, S., Nadiga, S., Thiaw, C., Wang, J., Wang, W., Zhang, Q., van den Dool, H. M., Pan, H.-L., Moorthi, S., Behringer, D., Stokes, D., Peña, M., Lord, S., White, G., Ebisuzaki, W., Peng, P., and Xie, P.: The NCEP climate forecast system, *J. Climate*, **19**, 3483–3517, 2006.

Saha, S., Moorthi, S., Pan, H-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, J., Hou, Y-T., Chuang, H-Y., Juang, H-M. H., Sela, J., Iredell, M., Treadon, R., Kleist, D., Delst, P. V., Keyser, D., Derber, J., Ek, M., Meng, J., Wei, H, Yang, R., Lord, S., Dool, H. V. D., Kumar, A., Wang, W., Long, C., Chelliah, M., Xue, Y., Huang, B., Schemm, J-K., Ebisuzaki, W., Lin, R., Xie, P., Chen, M., Zhou, S., Higgins, W., Zou, C-Z., Liu, Q., Chen, Y., Han, Y., Cucurull, L., Reynolds, R. W., Rutledge, G., and Goldberg, M.: The NCEP Climate Forecast System Reanalysis, *Bull. Amer. Meteor. Soc.*, **91**, 1015–1057, 2010.

Trenberth, K.E.: Global analyses from ECMWF and atlas of 1000 to 10 Mb circulation statistics, *NCAR Technical Note NCAR/TN-373+STR*, DOI: 10.5065/D6WM1BCN, 1992.

Uppala S. M., Kallberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Holm, E., Hoskins, B. J., Isaksen, L., Janssen, P. A. E. M., Jenne, R., McNally, A. P., Mahfouf, J. F., Morcrette, J. J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P. and Woollen, J.: The ERA-40 re-analysis, *Quart. J. Roy. Meteor. Soc.*, **131**, 2961–3012, 2005.

Wallace, J. M.: General circulation of the tropical lower stratosphere, *Rev. Geophys.*, **11**, 191–222, 1973.

Yoo, C., and Son, S.-W. : Modulation of the boreal wintertime Madden-Julian oscillation by the stratospheric quasi-biennial oscillation, Geophys. Res. Lett., 43, 1392–1398, doi:10.1002/2016GL067762, 2016.

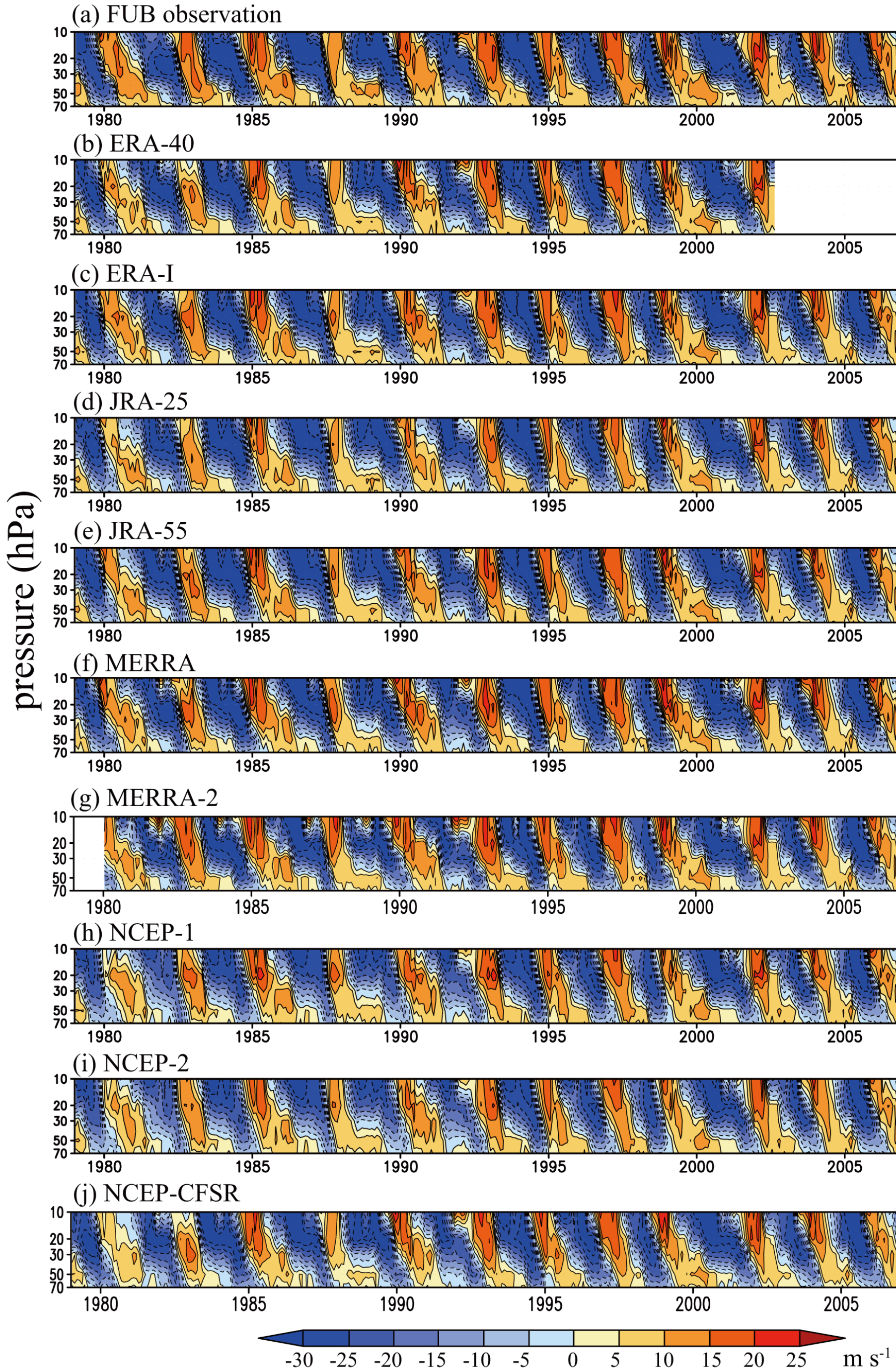


Figure 1. Time-height cross section of the FUB zonal wind and each reanalysis zonal wind over Singapore (104˚E, 1.4˚N) from 1979 to 2012. The color intervals are 5 m s^{-1}.

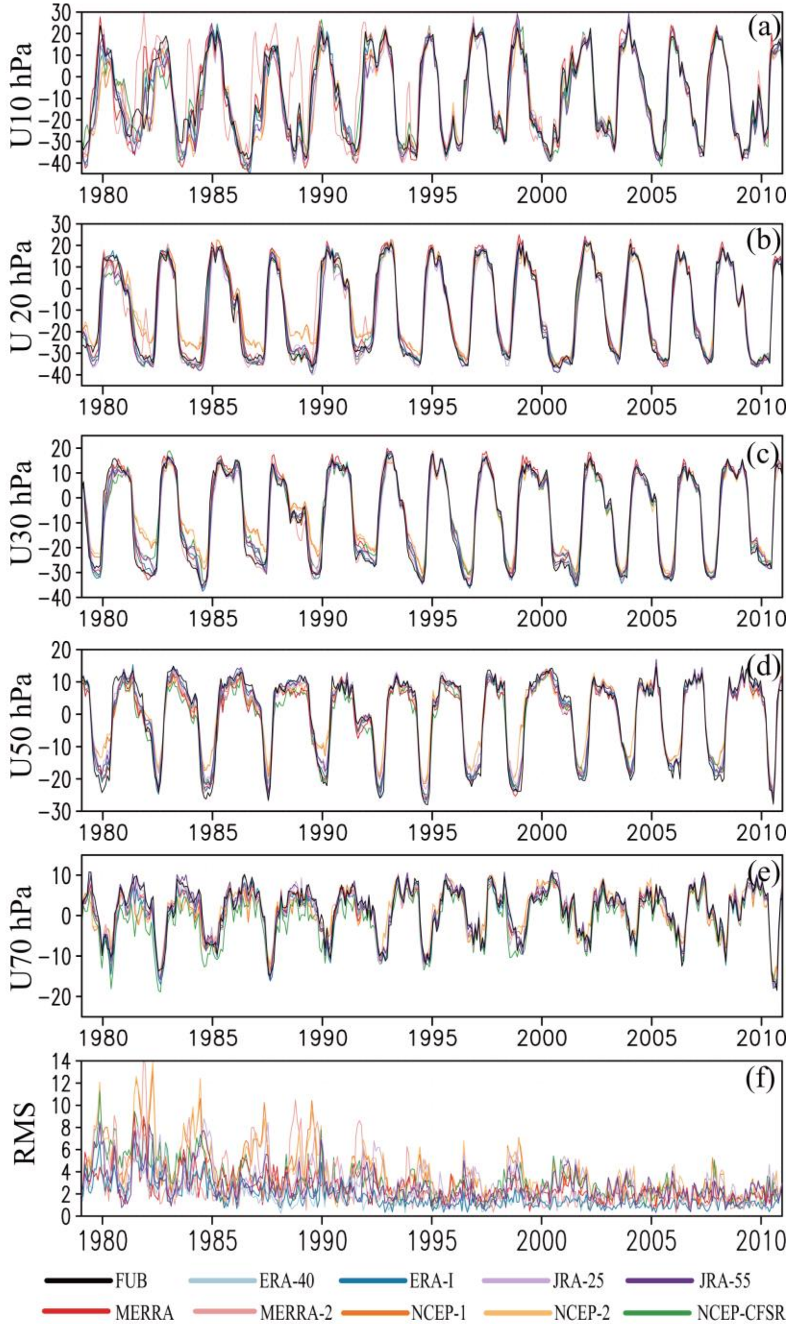


Figure 2. Time variations of the zonal wind over Singapore at (a) 10 hPa, (b) 20 hPa, (c) 30 hPa, (d) 50 hPa, and (e) 70 hPa for FUB observation (black), ERA-40 (light blue), ERA-I (blue), JRA-25 (light purple), JRA-55 (purple), MERRA (red), MERRA-2 (pink), NCEP-1 (orange), NCEP-2 (yellow) and NCEP-CFSR (green); (f) Root mean square differences between the FUB and each reanalysis zonal wind, averaged at 70–10 hPa.

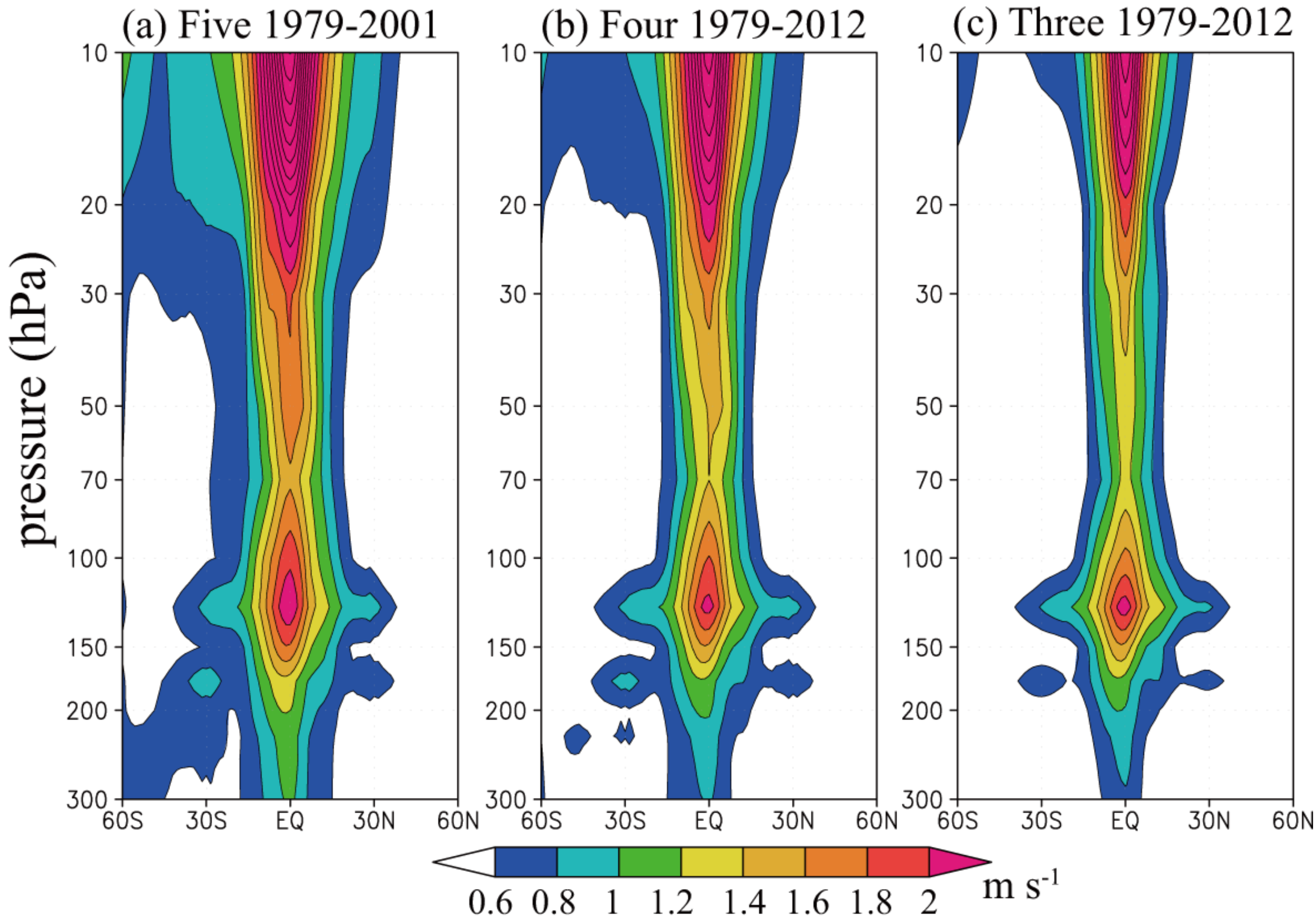
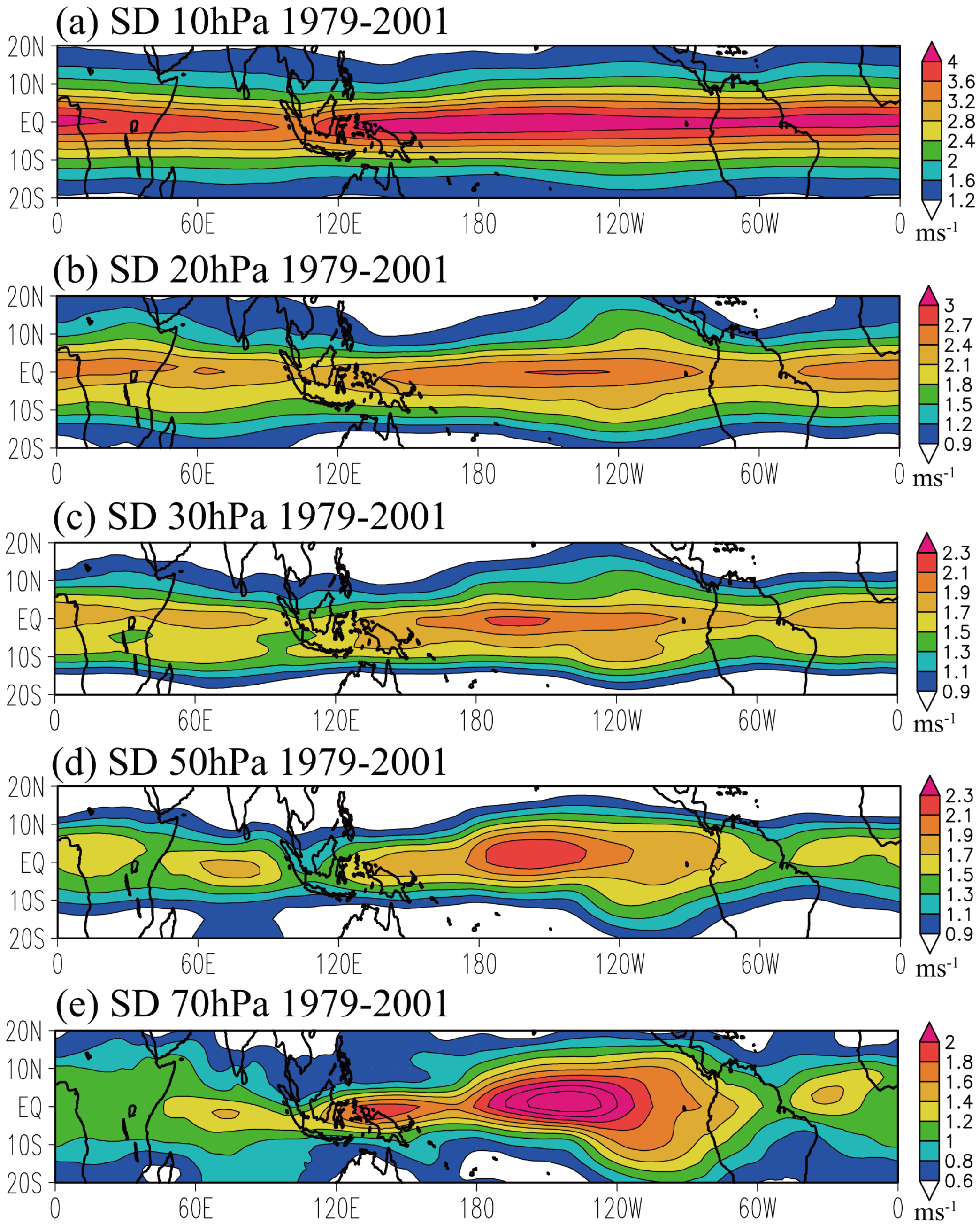
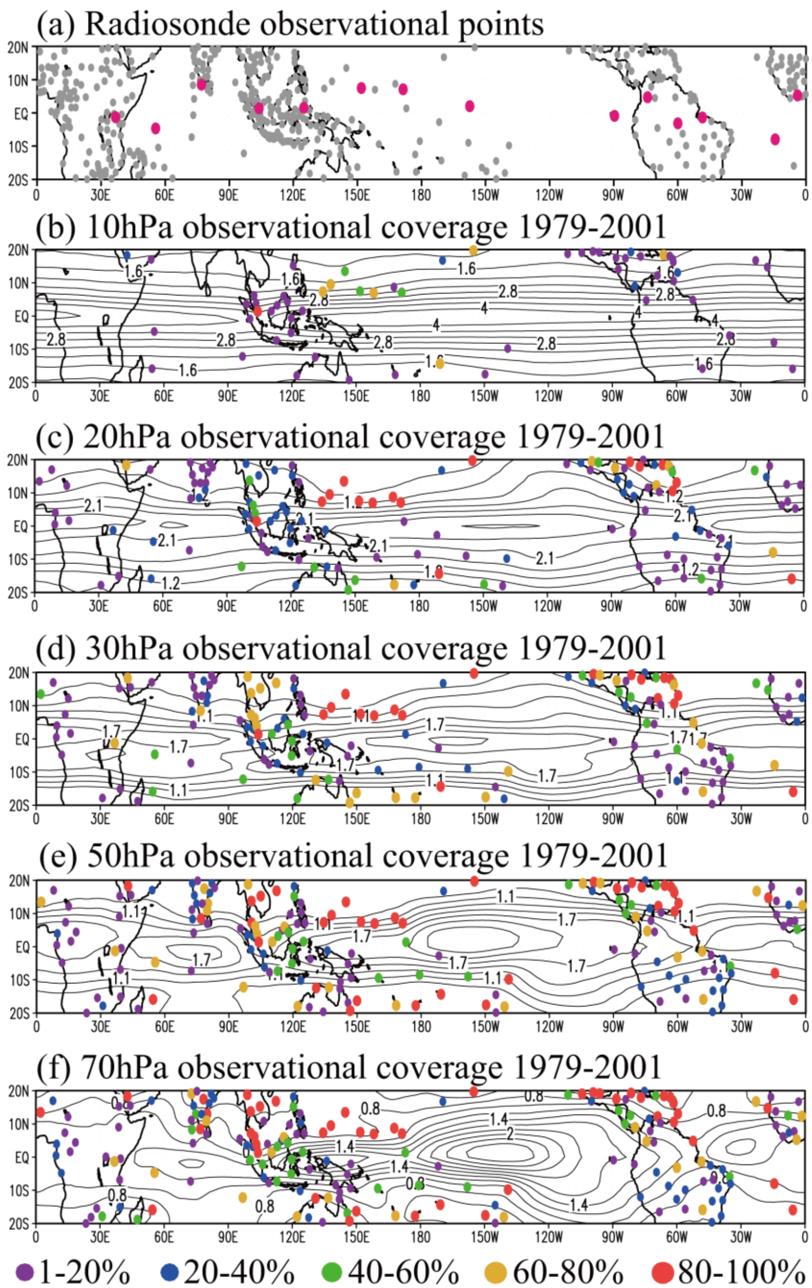


Figure 3. Latitude-height cross-sections of zonal mean and time mean of the standard deviation among (a) five reanalyses of ERA-40, ERA-I, JRA-25, JRA-55, and MERRA from 1979 to 2001, (b) four reanalysis of ERA-I, JRA-25, JRA-55, and MERRA from 1979 to 2012, and (c) three reanalysis of ERA-I, JRA-55, and MERRA from 1979 to 2012. The color intervals are 0.2 m s^{-1} and shaded with values larger than 0.6 m s^{-1}.

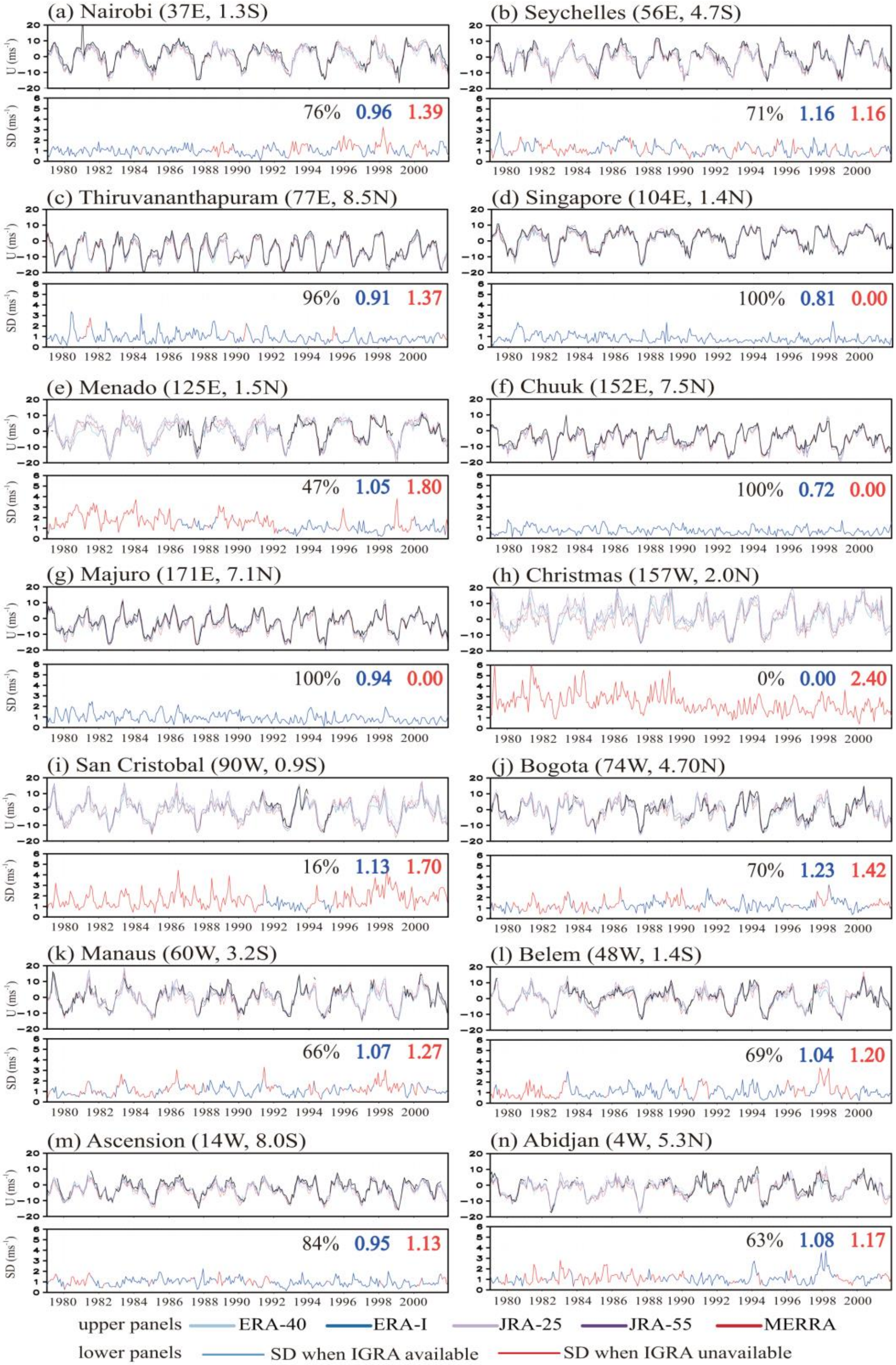
~~Figure 4. Vertical profiles of the standard deviation among the five reanalyses (ERA-40, ERA-I, JRA-25, JRA-55, and MERRA) at 10°S–10°N, averaged from 1979 to 2001. Black, blue, and red lines show the standard deviation of all, zonal mean, and eddy components, respectively.~~



~~Figure 5~~Figure 4. Horizontal maps of the standard deviation among the five reanalyses at (a) 10 hPa, (b) 20 hPa, (c) 30 hPa, (d) 50 hPa, and (e) 70 hPa: The color intervals are 0.4 m s^{-1} for (a), 0.3 m s^{-1} for (b), and 0.2 m s^{-1} for (c-e). Color ranges are different among these heights.



~~Figure 6~~Figure 5. (a) Locations of all IGRA stations in the tropical region (magenta dots indicate the locations of the stations shown in ~~Fig. 7~~Fig. 6). (b-f) IGRA stations with data coverage of (purple) 1–20%, (blue) 20–40%, (green) 40–60%, (yellow) 60–80% and (red) 80–100% at (b) 10 hPa, (c) 20 hPa, (d) 30 hPa, (e) 50 hPa, and (f) 70 hPa during 1979–2001. The contours show the standard deviation among reanalyses as shown in ~~Fig. 5~~Fig. 4. The contour intervals at each height are the same as in ~~Fig. 5~~Fig. 4.


(a) Nairobi (37E, 1.3S)
76% 0.96 1.39
(b) Seychelles (56E, 4.7S)
71% 1.16 1.16
(c) Thiruvananthapuram (77E, 8.5N)
96% 0.91 1.37
(d) Singapore (104E, 1.4N)
100% 0.81 0.00
(e) Menado (125E, 1.5N)
47% 1.05 1.80
(f) Chuuk (152E, 7.5N)
100% 0.72 0.00
(g) Majuro (171E, 7.1N)
100% 0.94 0.00
(h) Christmas (157W, 2.0N)
0% 0.00 2.40
(i) San Cristobal (90W, 0.9S)
16% 1.13 1.70
(j) Bogota (74W, 4.70N)
70% 1.23 1.42
(k) Manaus (60W, 3.2S)
66% 1.07 1.27
(l) Belem (48W, 1.4S)
69% 1.04 1.20
(m) Ascension (14W, 8.0S)
84% 0.95 1.13
(n) Abidjan (4W, 5.3N)
63% 1.08 1.17
U (ms⁻¹)
SD (ms⁻¹)
1980 1982 1984 1986 1988 1990 1992 1994 1996 1998 2000
upper panels ERA-40 ERA-I JRA-25 JRA-55 MERRA
lower panels SD when IGRA available SD when IGRA unavailable

~~Figure 7~~Figure 6. Upper panels show observed zonal wind (black) at each station and reanalysis zonal wind (color) at 70 hPa interpolated to each observational point from 1979 to 2001. Lower panels depict standard deviation among the five reanalyses at 70 hPa. Blue lines show the standard deviation during times when monthly-mean radiosonde data is available and red lines show the standard deviation during times when they are not. In the upper right corner of each bottom panel, the total data coverage (%) at each station and the average value of the standard deviation during periods with (blue number) and without (red number) radiosonde data are shown.

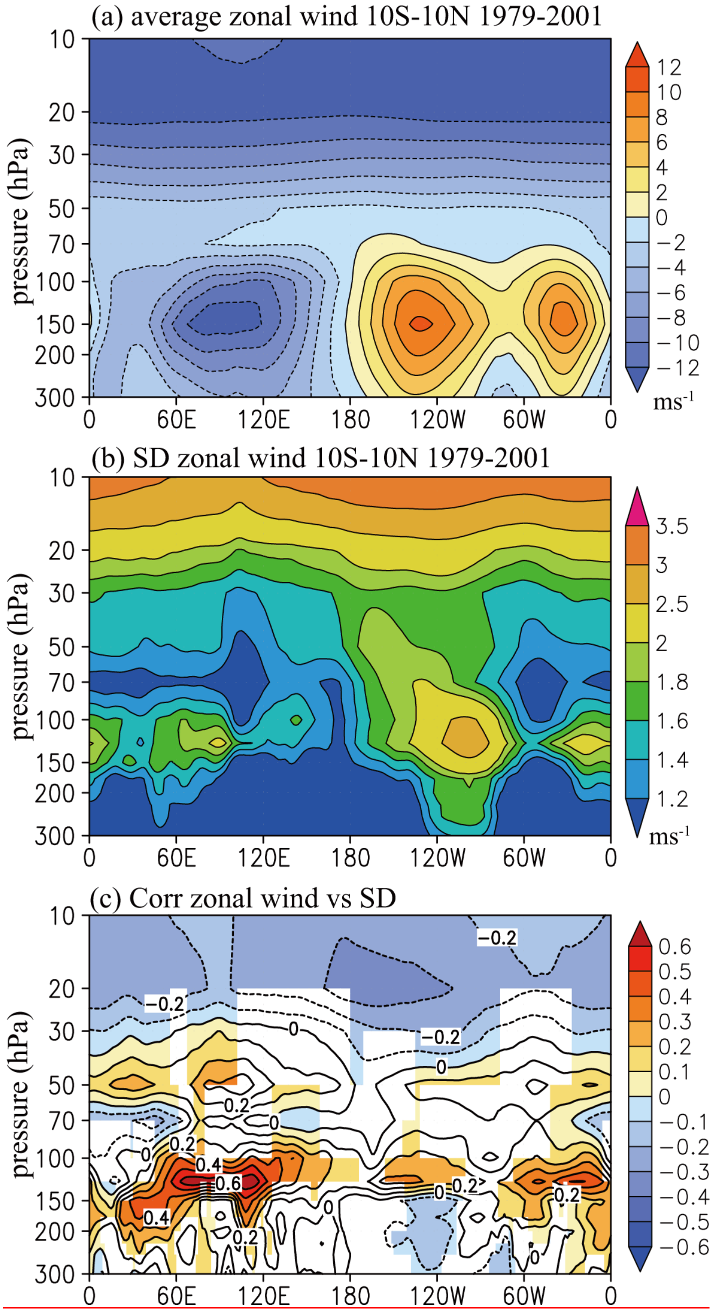


Figure 7. Longitude-height cross section of (a) averaged zonal wind in the five reanalyses, and (b) the standard deviation among reanalysis and (c) the correlation between absolute value of zonal wind and the standard deviation in 10°N–10°S averaged from 1979 to 2001. The color intervals are (a) 2 m s^{-1}, (b) 0.2 m s^{-1} for values less than 2 m s^{-1} and 0.5 m s^{-1} for values more than 2 m s^{-1}, (c) 0.1, with statistical significance of ≥95% colored.

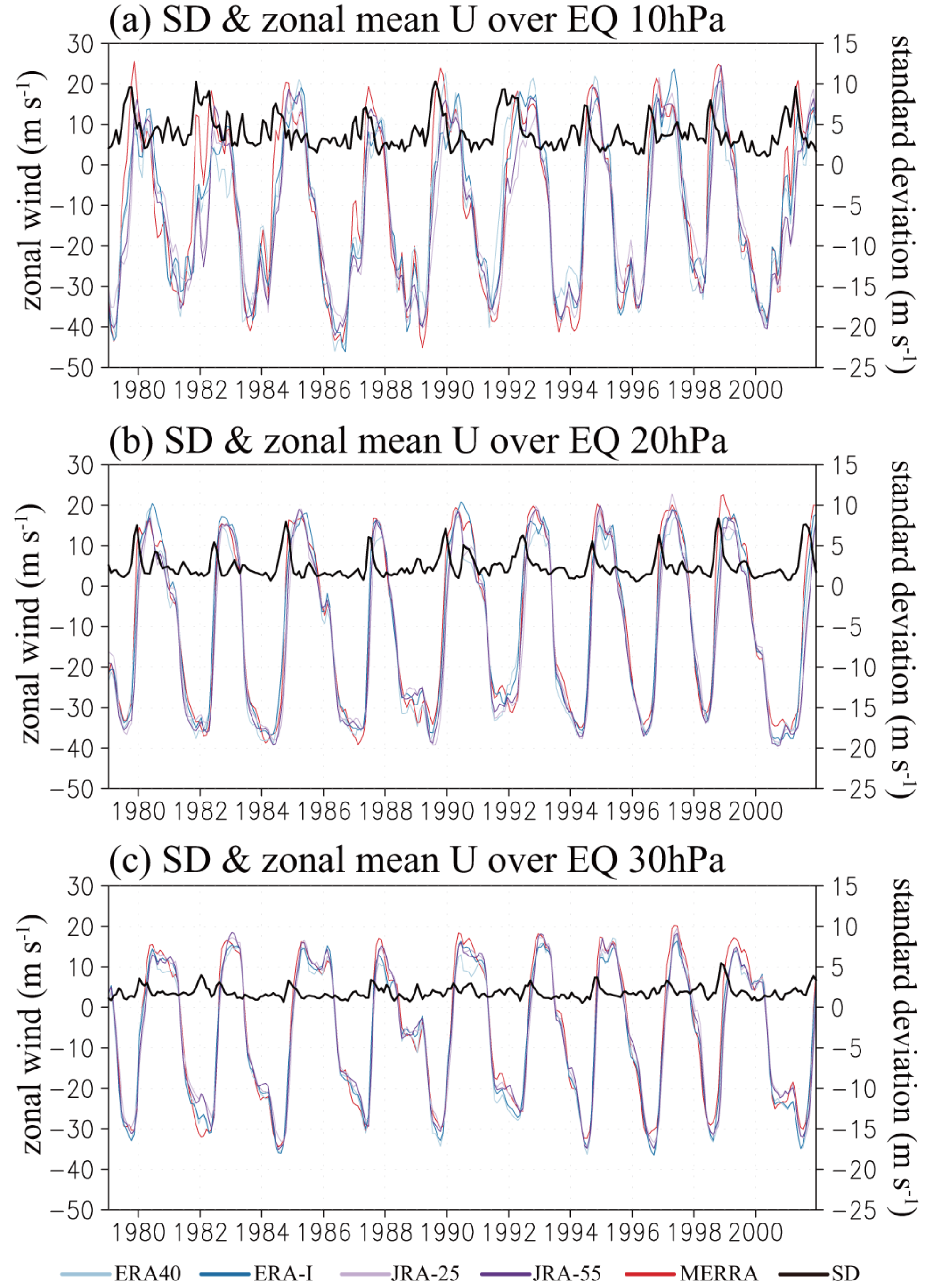


Figure 8. Zonal mean equatorial zonal wind in each reanalysis (colors) and zonally averaged equatorial standard deviation among reanalyses (black) at (a) 10 hPa, (b) 20 hPa, and (c) 30 hPa.

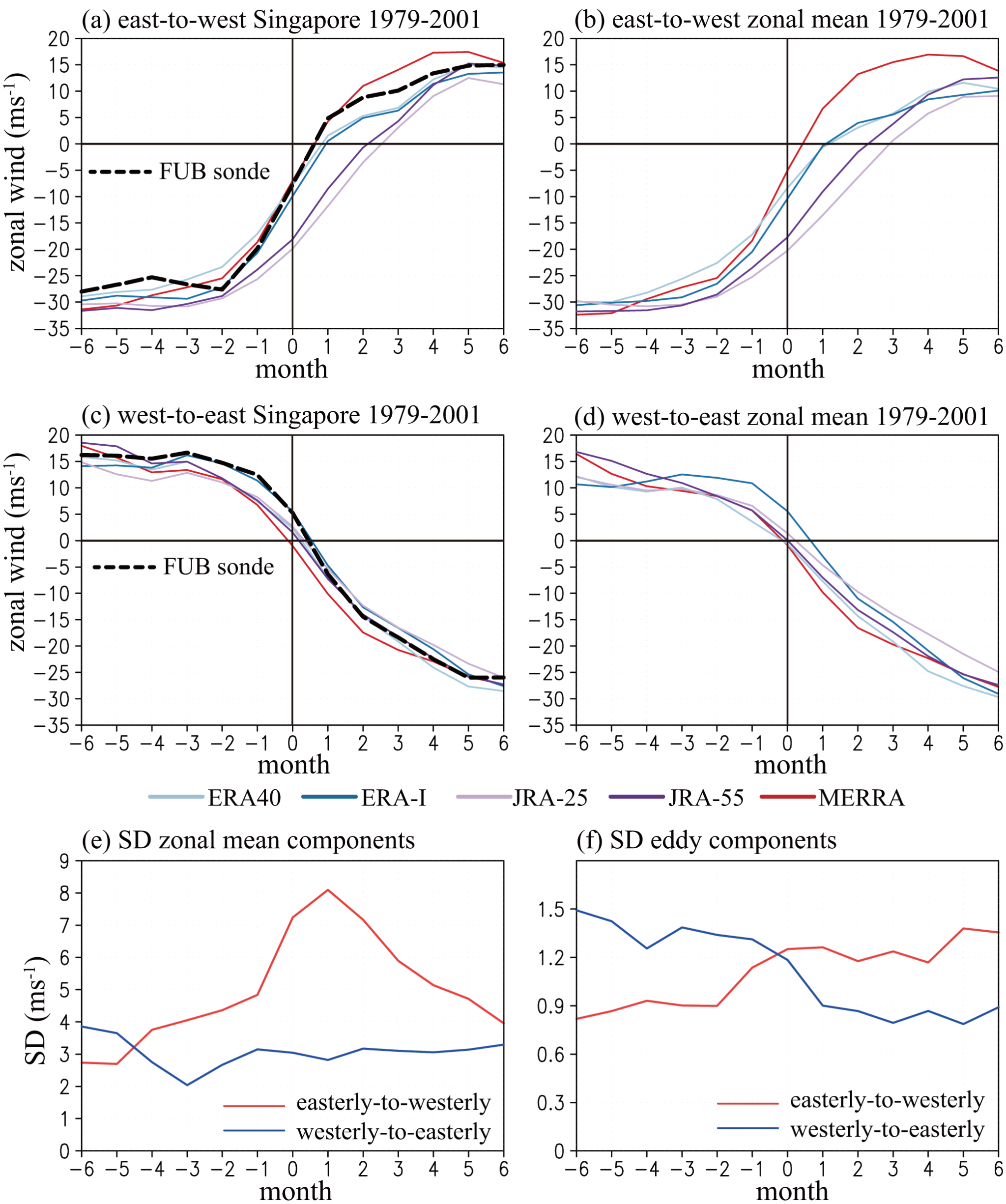


Figure 9. Top: Composite of the QBO in the zonal wind at 10 hPa during 1979 to 2001 where month 0 corresponds to (a, b) the easterly-to-westerly transition of the FUB zonal wind at 10 hPa and to (c, d) the westerly-to-easterly transition. Color lines show each reanalysis zonal wind, and black dashed lines depict FUB zonal wind at Singapore. Results for (a, c) the zonal wind at Singapore and for (b, d) the zonal mean equatorial zonal wind. Bottom: Composite of the zonal mean equatorial standard deviation due to (e) zonal mean and (f) eddy components during (red) easterly-to-westerly and (blue) westerly-to-easterly transitions at 10 hPa. Note the ordinate axes are different between (e) and (f).

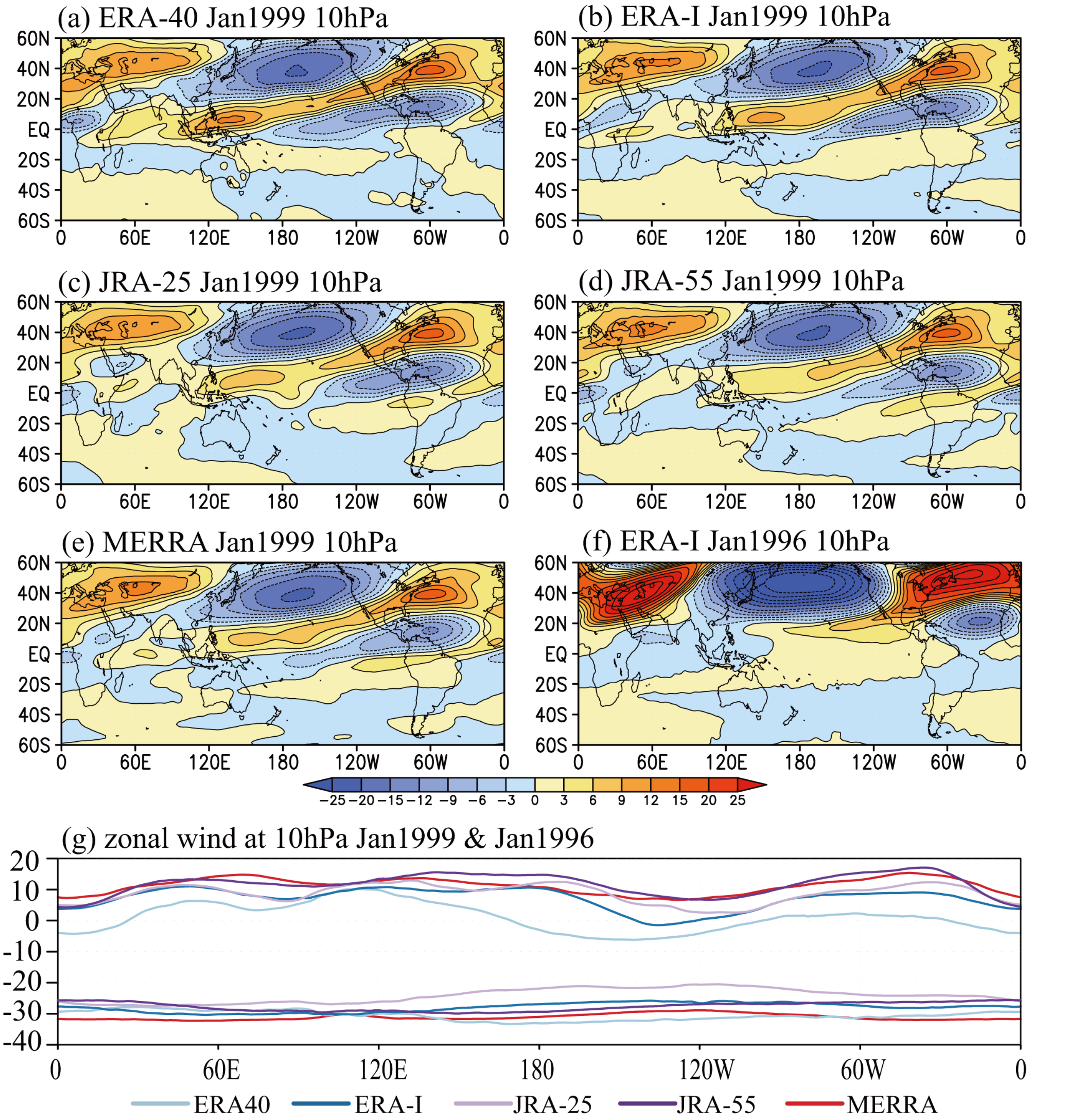


Figure 10. Deviation of the zonal wind values from the zonal mean (u') at 10 hPa of (a-e) five reanalyses in January 1999 during westerly phase and (f) of the ERA-I in January 1996 during easterly phase of the QBO. The color interval is ±3, 6, 9, 12, 15, 20, or 25 m s^{-1}. (g) Longitudinal variations of the zonal wind of each reanalysis at 10 hPa over the equator in January 1999 and January 1996.

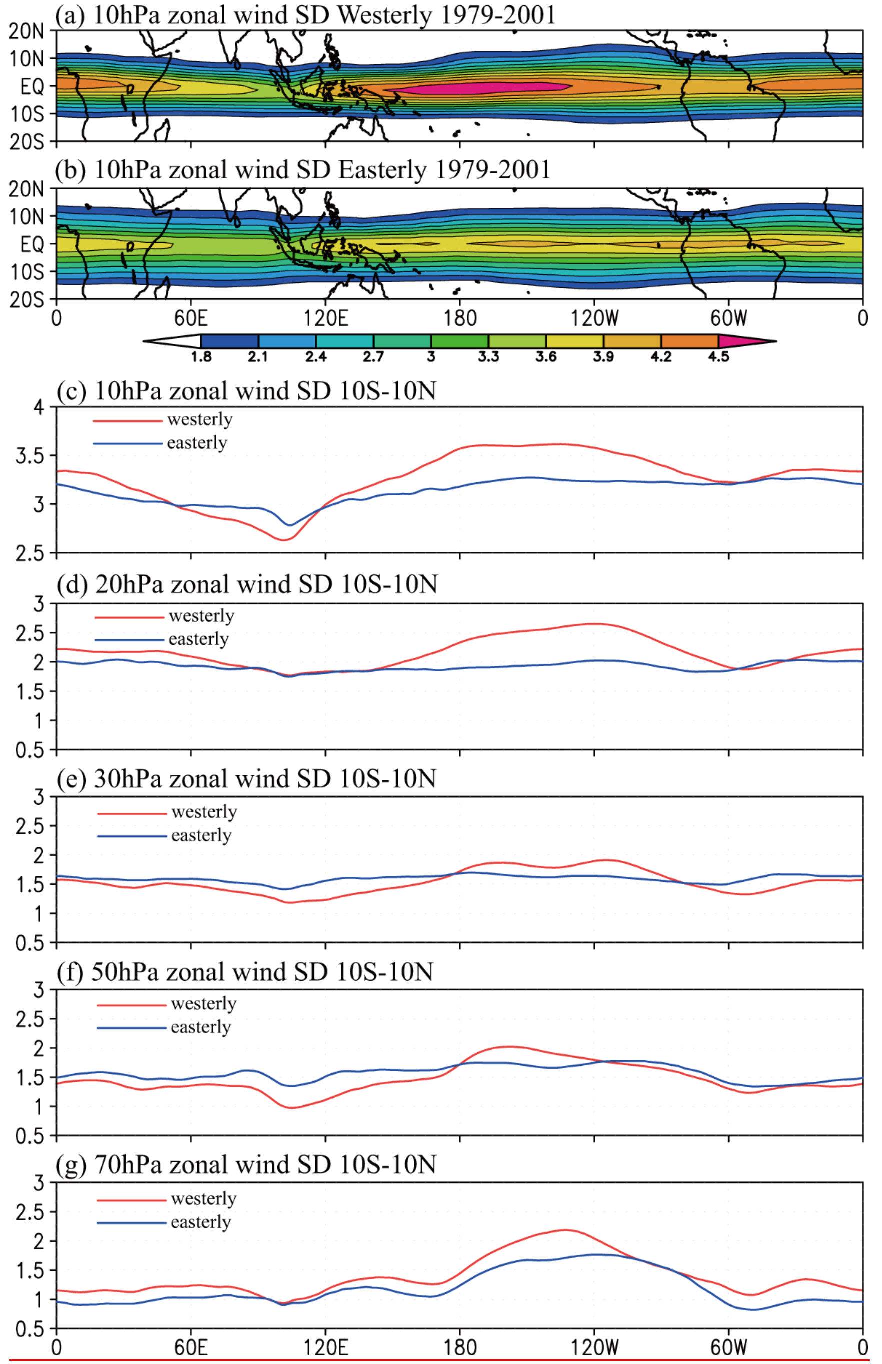
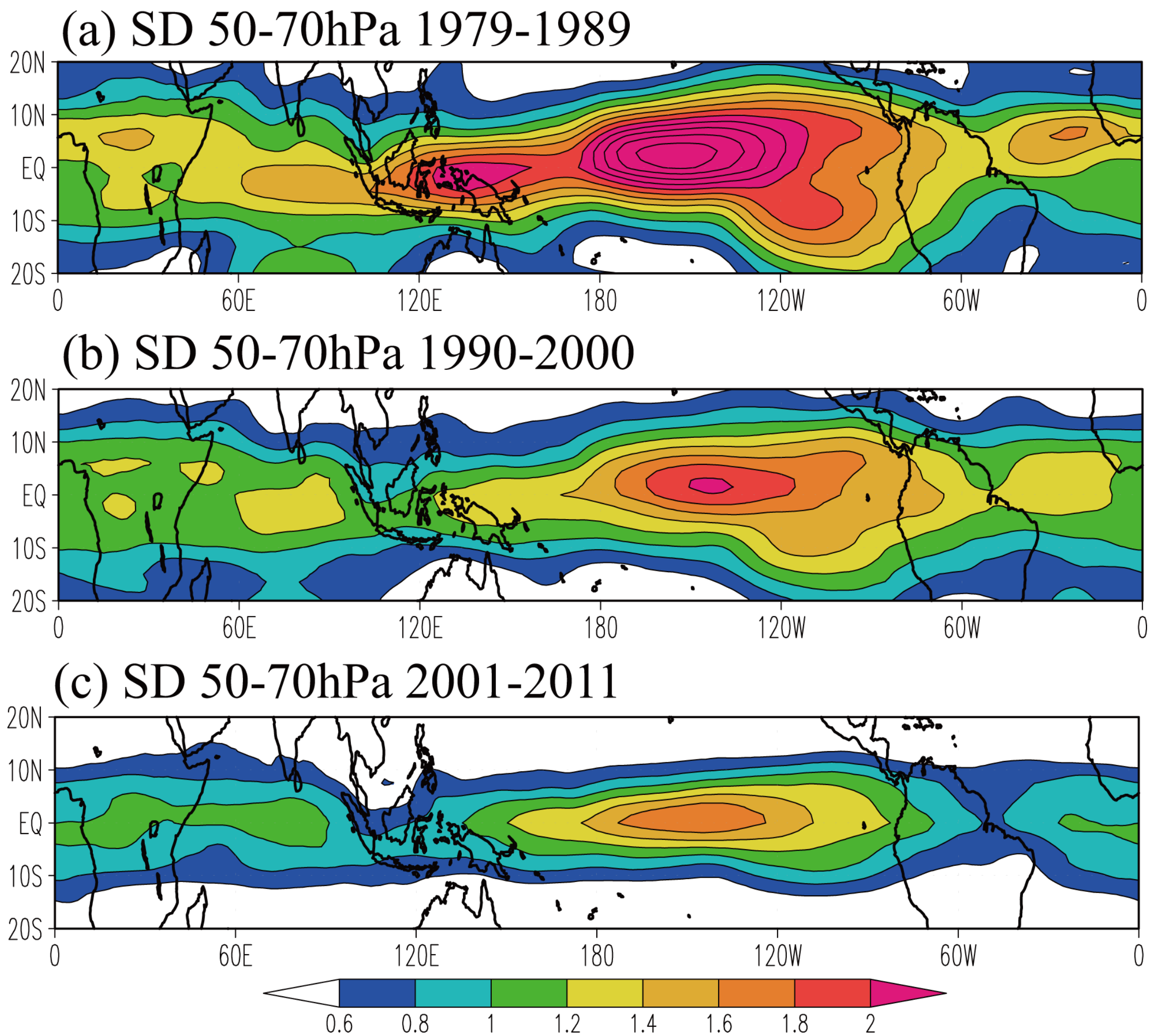
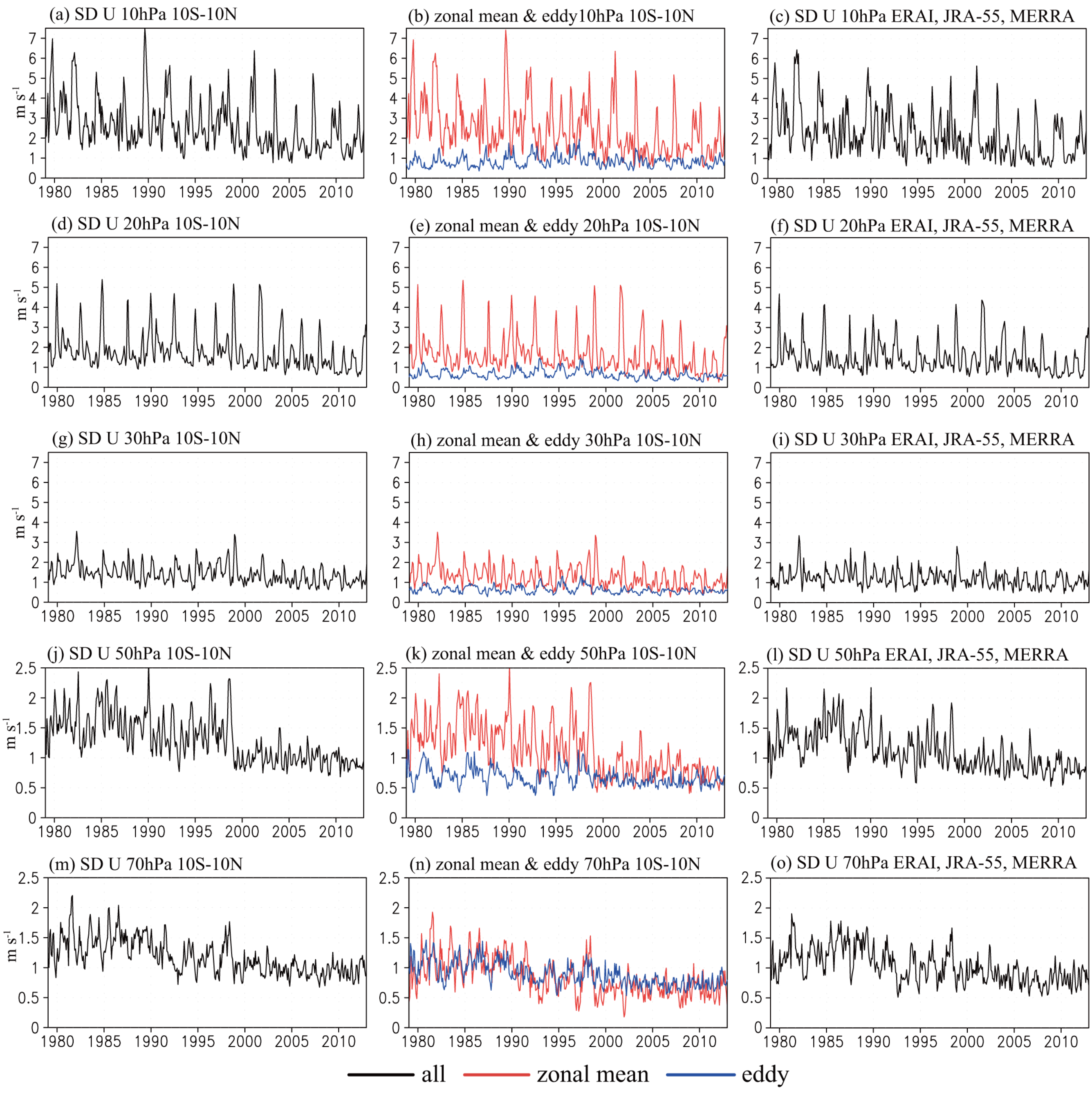


Figure 11. Horizontal distributions of the standard deviation among reanalyses at 10 hPa during (a) westerly and (b) easterly phase of the QBO averaged in the period 1979–2001. The color interval is 0.3 m s^{-1}. ~~(c)~~ Longitudinal variations of the standard deviation at 10˚S–10˚N in (red) westerly and (blue) easterly phase at (c) 10, (d) 20, (e) 30, (f) 50, and (g) 70 hPa.

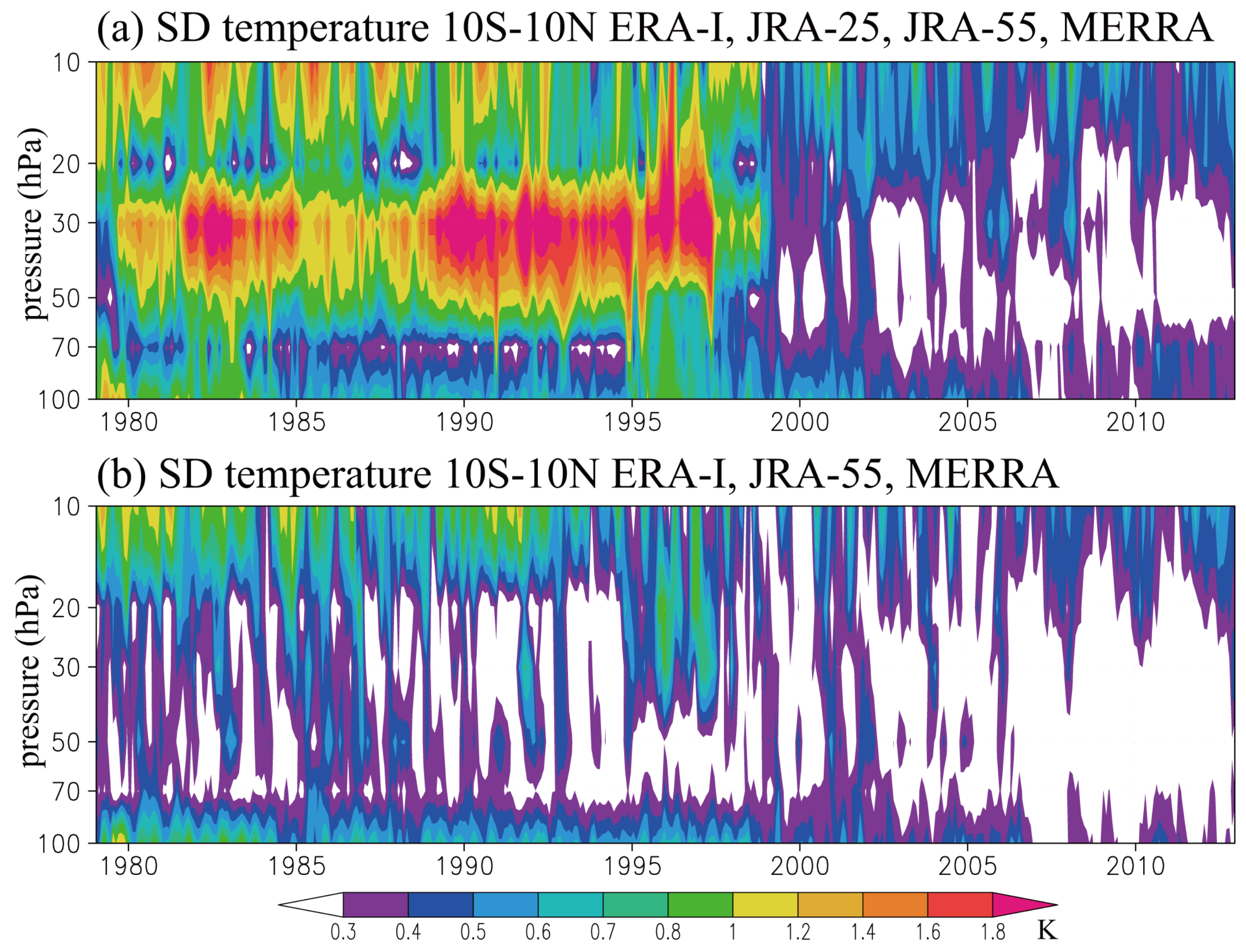
~~Figure 13. Longitude-height cross section of (a) averaged zonal wind in the five reanalyses, and (b) the standard deviation among reanalysis in 10°N–10°S averaged from 1979 to 2001. The color intervals are 2 m s^{-1} for (a) and 0.2 m s^{-1} with values less than 2 m s^{-1} and for (b) 0.5 m s^{-1} with values more than 2 m s^{-1}.~~



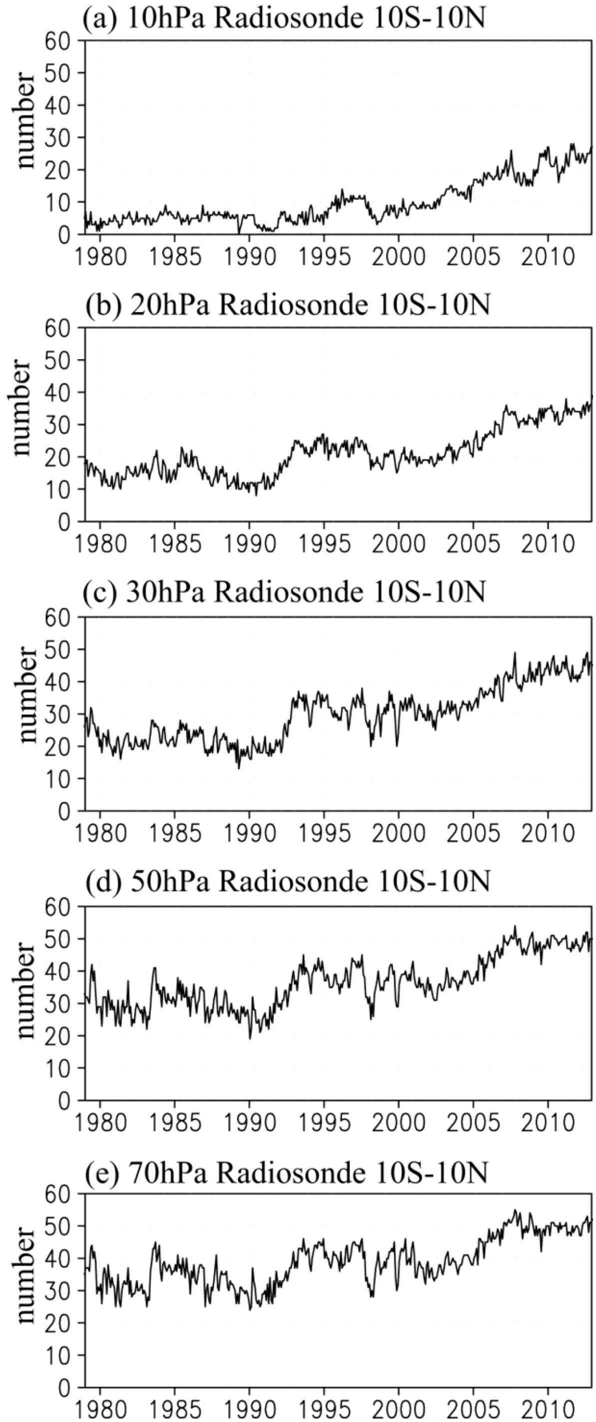
~~Figure 14~~Figure 12. Standard deviation among four reanalyses (ERA-I, JRA-25, JRA-55, and MERRA) at 50-70 hPa for each 11-year mean: (a) 1979–1989, (b) 1990–2000, and (c) 2001–2011. The color intervals are 0.2 m s^{-1}.



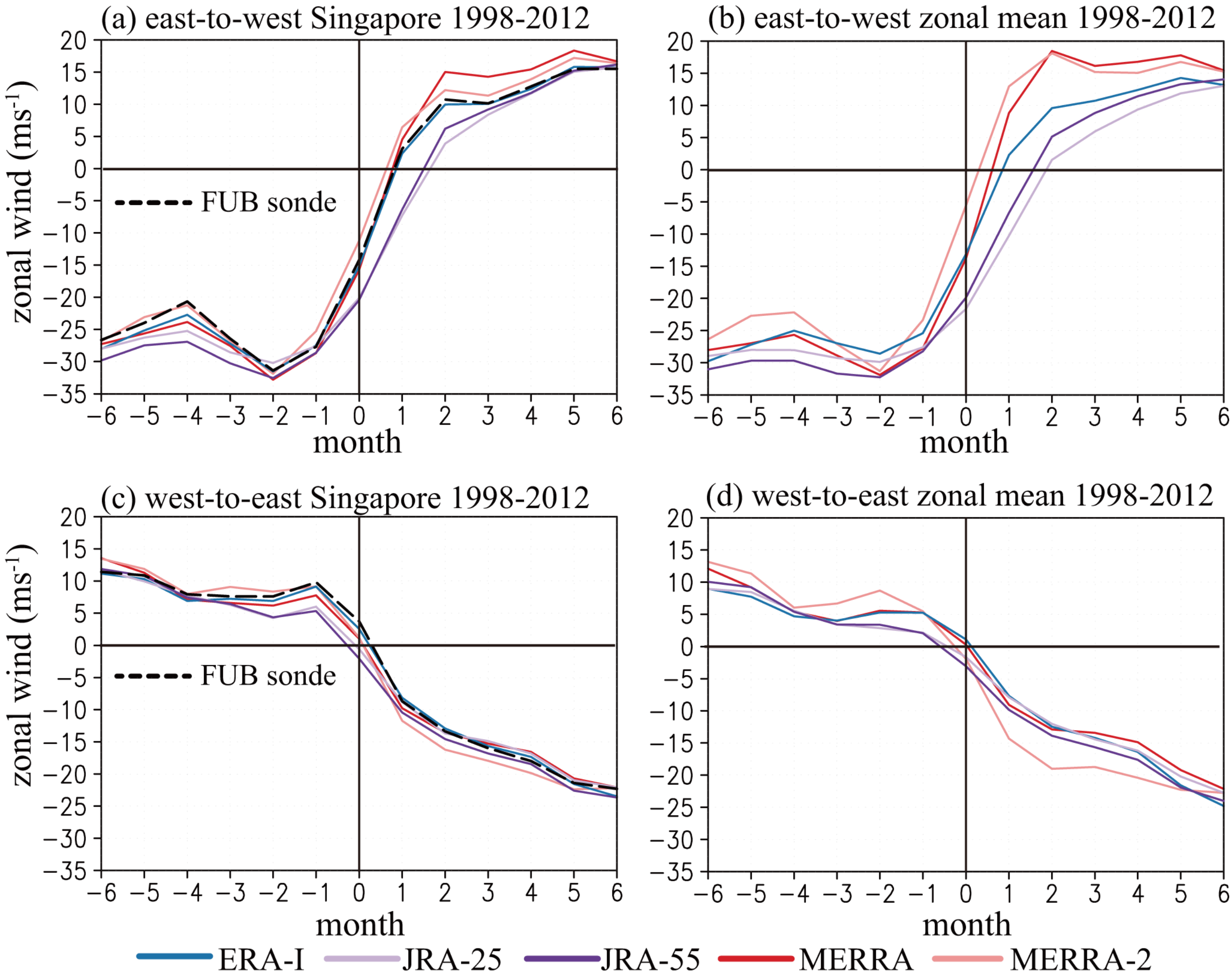
~~Figure 15~~Figure 13. Time variations of zonal mean standard deviation of (black) all, (red) zonal mean, and (blue) eddy components among (left and middle) four reanalyses (ERA-I, JRA-25, JRA-55, and MERRA), and among (right) three reanalyses (ERA-I, JRA-55, and MERRA) from 1979 to 2012 at 10 hPa, 20 hPa, 30 hPa, 50 hPa, and 70 hPa in 10˚S-10˚N.



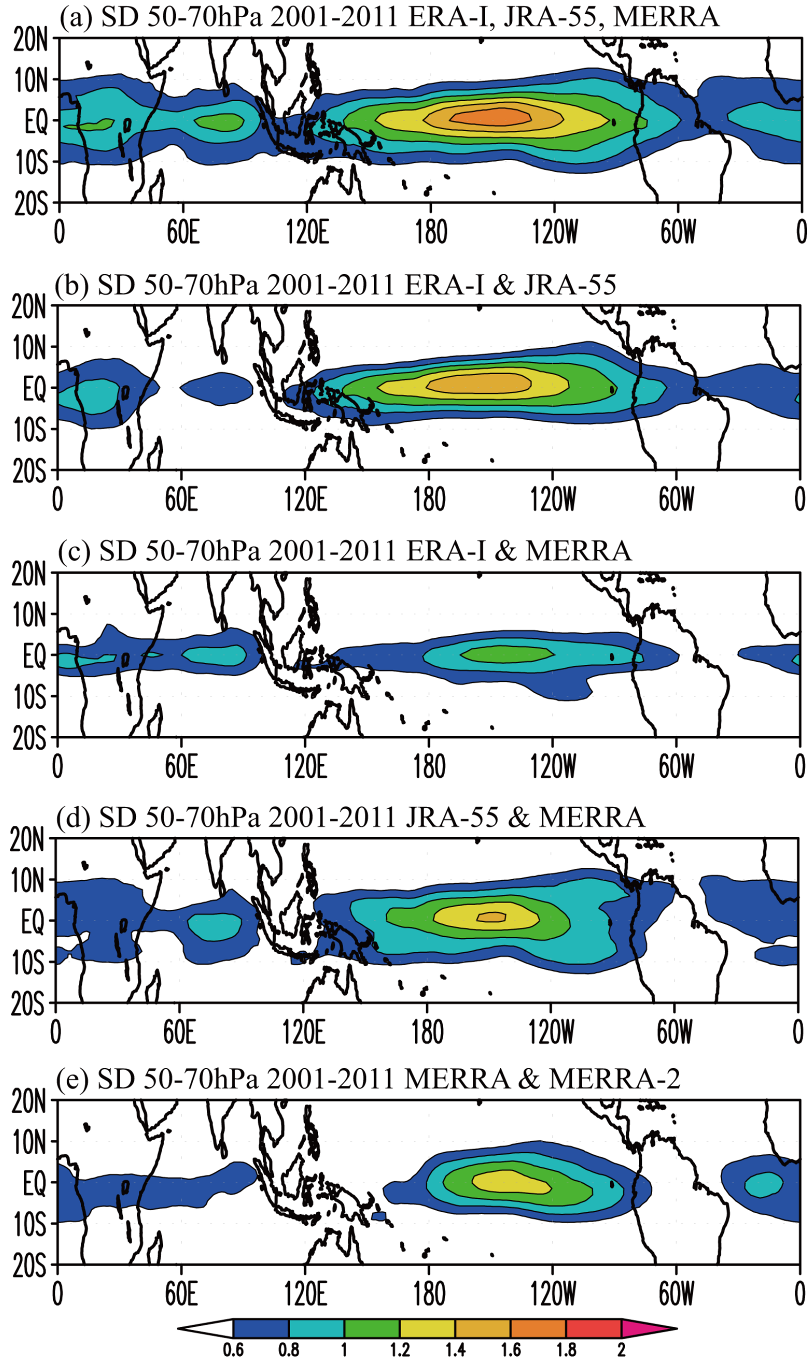
~~Figure 16~~Figure 14. Time-height cross sections of zonal mean standard deviation of temperature among (a) four reanalyses (ERA-I, JRA-25, JRA-55, and MERRA), and (b) three reanalysis (ERA-I, JRA-55, and MERRA) at 10°S–10°N from 1979 to 2012. The color intervals are 0.1 K for values less than 0.8K and 0.2 K for values more than 0.8K.



~~Figure 17~~Figure 15. Time variation in number of radiosonde observations at 10°S–10°N at (a) 10 hPa, (b) 20 hPa, (c) 30 hPa, (d) 50 hPa, and (e) 70 hPa from 1979 to 2012.



~~Figure 18~~Figure 16. Same as Figs. 9a-d but during 1998 to 2012, excluding ERA-40 and including MERRA-2.



~~Figure 19~~Figure 17. Same as ~~Fig. 14~~Fig. 12c but for the standard deviation among (a) three reanalysis (ERA-I, JRA-55, and MERRA), (b) between ERA-I and JRA-55, (c) between ERA-I and MERRA, (d) between JRA-55 and MERRA, and (e) between MERRA and MERRA-2.