Interactive comment on “Defining metrics of the Quasi-Biennial Oscillation in global climate models” by Verena Schenzinger et al.

Anonymous Referee #1

The Quasi-Biennial Oscillation is one of the most important modes of variability in the atmosphere and it is to an increasing extent included in climate models and CCMs. The present paper defines a set of metrics for the QBO and compares these metrics for a set of models and reanalyses.
The subject is important and the paper is well written. However I have a couple of major considerations that the authors should consider before I can recommend that the paper is accepted.
Major comments:
1) I miss some motivations for the chosen metrics and the way they are defined.

Added “These metrics were defined to be as simple as possible, yet meaningful in characterising the QBO morphologically. For robust and simple assessment of the QBO in models and observations, this study focusses on the large-scale morphology of the QBO rather than those (small-scale) dynamical processes involved in maintaining it.” (p. 2, l. 18-20)

For example, some metrics are defined from the Fourier filtered time-series while others seems to be defined from the raw zonal mean zonal wind. What would the difference be if the metrics were calculated from the real data without the Fourier filtering?

This is not the case. Metrics are either defined from the raw zonal wind or the spectrum. No filtering is applied. The section describing the metrics definitions has been rewritten for clarification (p. 3, l. 3 - p. 4, l. 10).

Why are the mean period defined from calculating zero-crossings in the raw zonal mean zonal wind and not from the spectrum?

Using the zonal wind is a more intuitive and accurate way to define the period and give an error estimate. Looking at the Fourier spectrum of the ERA-Interim reanalysis (Figure 2, left panel), the periods calculated from the timeseries compare well to the broad spectral peak.

In many studies a filtering based on the leading principal components are used (e.g., Wallace 1993, JAS 50, 1751-1762) making it possible to obtain a well defined a phasespeed. This possibility is not even mentioned in the paper.

One of the aims was to define metrics as simple and intuitively as possible - PC analysis introduces an additional analysis step that would need further justification; calculating periods from the first two principal components for the observations gives the same result as from the wind within the standard deviation interval (28.0 ± 3.6 vs. 28.2 ± 4.4 months). This was added as a footnote in the metrics description (p. 3)

I also wonder why there is no metric related to the wave-forcing of the QBO included.

The metrics are primarily for phenomenological assessment and also the necessary data were not available for most of the models. Added “These metrics were defined to be as simple as possible, yet meaningful in characterising the QBO morphologically. For robust and simple assessment of the QBO in models and observations, this study focusses on the large-scale morphology of the QBO rather than those (small-scale) dynamical processes involved in maintaining it.” (p. 2, l. 18-20) to clarify.

The metrics could also be somewhat more detailed described in the text. Even when the caption to Fig. 2 is included the definitions are very densely described.

Agreed - the definition of metrics has been expanded to an algorithmic description (p. 3, l. 3 - p. 4, l. 10).

For example,
how is the cut-off frequencies of the QBO in the spectrum around two years actually determined?

Cut-off frequencies are calculated from the minimum and maximum period; description has been changed to clarify (p. 3, l. 22).

2) The second half of the paper deals with "model performance" and metrics calculated for the models are compared to those of observations. But there is almost no attempt to address the statistical uncertainty (in e.g. Table 3). Given the relatively few QBO events on record this is an important part of the analysis both for the comparisons in this paper and in general.

Added paragraph “Error estimation” (p. 4, l. 15-30) and numbers in tables 3 and 4.

The monthly QBO data can because of the oscillatory nature not be modelled by simple processes such as white noise or AR1 models. Christiansen 2010 (JCLIM, 23, 3953-3966) demonstrates one way to overcome this with a Monte Carlo method.

Thank you for the suggestion. The method was included in error estimation for the minimum/maximum period and the Fourier amplitude (p. 4, l. 18-25). Outside the area that is dominated by the QBO (above 10hPa, below 70hPa, further away from the equator) the method unfortunately cannot be used as no clear QBO cycle can be defined.

Anyway, the authors should address this problem and provide uncertainty intervals for the numbers in the tables.

Uncertainties are now included in tables 3 and 4.

Minor comments:
Lines 11, 16: What is meant by "easterly/westerly shear zones"?. Here it seems just to be the easterly/westerly phases the zonal mean wind.
Changed wording to “phase” where appropriate.

Fig. 3 and 5: Are the profiles in Fig. 3 for one single QBO event? Yes. The caption states that these are from the 1964-1966 cycle. And are the mean and standard deviations shown in Fig.5 then taken over all QBO events? Yes. In the metrics description, it is defined as “The mean of the descent rates between 10 and 70hPa is calculated separately for the two shear zones as the mean over all values for a descending easterly/westerly.” (p. 4, l. 10)

Perhaps more details about the models could be included in Table 1 regarding the parameterizations of the orographic/non-orographic gravity waves. Relevant references are included

in the table.
Table 2: Is "mth" in the unit for decent rates the same as "months"? Changed mth->month

Thank you for your review.

Anonymous Referee #2

This study evaluates the QBO as represented in recent climate models, using the small number of CMIP5 and CCMVal models that represent the QBO. The main point of the study is establish the set of metrics that are used here to characterize the QBO, and the authors advocate that these metrics be used in future multi-model comparisons such as those expected from the SPARC QBOi activity. An interesting finding is that the models, on average, have QBOs that are shifted upward and are meridionally too narrow, in comparison to reanalyses.
The proposed metrics are potentially a timely and useful contribution. However I have some issues with the way in which they are presented:
1. The method for calculating the metrics should be presented in a crystal-clear, algorithmic fashion. Since the point is for future studies to repeat these calculations on different models and/or reanalyses, it needs to be very clear how to do this. I don't think the description of the calculations is sufficiently clear in the present draft. Please see detailed comments in the line-by-line remarks, below.
Thank you for this remark. The definition of the metrics has been changed to an algorithmic description.
(p. 3, l. 3 - p. 4, l. 10).

2. The metrics are presented as-is, with virtually nothing being said on why these choices were made and not others. For example, other ways of defining the QBO amplitude have appeared in the literature, such as Baldwin and Gray 2005. It would be useful for the authors to make the case as to why they settled on these particular choices. Otherwise I speculate that later authors might choose different metrics to characterize the QBO, if this paper hasn't convinced them that the choices made here are well founded. For example, why not just simply use the RMS monthly-mean zonalmean wind amplitude at a set of standard pressure levels as the measure of QBO amplitude?

Metrics have been chosen for simplicity and conciseness - having an amplitude definition at a set of levels is more complicated (more numbers) than defining it at one particular level that has been chosen as the level where the QBO is strongest.
Added "These metrics were defined to be as simple as possible, yet meaningful in characterising the QBO morphologically." (p. 2, l. 18-19) to clarify the general approach to choosing the metrics.

3. The metrics in Tables 3 and 4 have no uncertainty estimates associated with them and I see no reason for that omission. The results are mostly given to three significant figures but there is no sense of how meaningful this precision is. Table 6 does give estimates, associated with the multi-model ensemble spread. But for single models (and reanalyses), shouldn't it be possible to give uncertainties based on the internal variability? That is, the variation between QBO cycles.

Added paragraph "Error estimations" (p. 4, l. 15-30) and uncertainties in tables 3 and 4.
Unfortunately not all numbers can be derived from variations between cycles; an alternative method has been applied to those (p. 4, l. 18-25).

Based on these issues, and other detailed comments below, I recommend major revisions.
Some other suggestions:
1. Plots for individual models would be useful as supplemental material. For example you could make Fig. 4 for each model individually.
As requested, the equivalent plots for the individual models were added to supplementary material.

2. Amplitude of the QBO in temperature at the tropical tropopause would be a useful metric. You would need to define the tropopause, but perhaps even just providing the amplitude at 100 hPa would be a simple and useful way to do it.
As the models have strong deficiencies to represent the dynamical QBO close enough to the tropopause, it is questionable how useful this metric would be. Further, temperature amplitudes at the tropopause are influenced by many other regional factors (e.g. ENSO), making it harder to attribute a certain variability to the QBO.
To address this comment, the "Depth" measurement for temperature, defined analogous to the "Depth" from the wind field, has been introduced.

3. It might be useful to state, in your discussion section, what interesting properties of the QBO are not captured by these metrics. For example some characterization of the zonal momentum budget would be interesting. I'm not suggesting the paper needs to include that, but it would good to state why it doesn't. Data not available in the CMIP5 archive? A desire for simplicity? The metrics are defined for simplicity and objectivity and only deal with the morphology of the QBO. Further, data is not available for all models in the CMIP5 archive and there is little observational reference.
Added " For robust and simple assessment of the QBO in models and observations, this study focusses on the large-scale morphology of the QBO rather than those (small-scale) dynamical processes involved in maintaining it." (p. 2, l. 19-20)

4. Histograms showing the distribution of QBO period in each model could be useful (a multi-panel plot, one panel per model). Fig 6 is useful, but the models might show interesting variations amongst themselves. It would show whether some models tend to be more synchronized with the annual cycle than others.
Added to supplementary material.

Comments/suggestions by page and line number:
Page 1
6: "QBO like" –> "QBO-like" Done.
12: "ERA-Interim" –> "ERA-Interim reanalysis" Done.
16: insert "known" after "repeatable" Done.
16: ", beyond" –> " outside of" Done.
17: From Osprey et al and Newman et al I think we've learned that whether the QBO remains as regular in present-day climate is also an outstanding question! Of course. Sentence changed to "Whether or not the QBO remains as regular in the present-day climate and under future climate change is an outstanding question".
Page 2
2-3: Join these two paragraphs together, since they both describe the basic QBO theory. Done.
7: insert "reasonably" before "realistic" Done.
8: "Follow on" –> "Follow-on" Done.
8: delete "signal" Done.
9-10: "depends, amongst others," –> "is affected by" Done.
14: Four out of thirty sounds pretty bad, but on the other hand many of these models

might have poor stratospheres in general, with model lids below the stratopause. Do you have an estimate of how many of the CMIP5 models can be regarded as "stratosphere-resolving" but still don't produce a QBO?

Ten models are stratosphere-resolving and include non-orographic gravity waves (Charlton-Perez, A. J., et al. *(2013), On the lack of stratospheric dynamical variability in low-top versions of the CMIP5 models*, J. Geophys. Res. Atmos., 118, 2494–2505), so should be able to produce a QBO. Added a footnote on page 2.

17: "aims" –> "aim", "are" –> "is" Done.

20: "An additional purpose is to provide" –> "The purpose is to provide" Done.

21: "the future QBO simulations" –> "new QBO-resolving" (so as not to suggest that only future projections are of interest) Done.

29: "Merra" –> "MERRA" Done.

Page 3

1-2: Suggest deleting this first sentence, it doesn't really add anything. You might instead start this section by introducing Figure 1, since otherwise the figure is first introduced in parentheses near the end of the paragraph, which is easily missed. Agreed and done.

4: "was quickly established": not sure what you're referring to here. A previous project comparing QBOs in different models? Changed to "The most obvious one is the mean period;"

5-6: "a typical oscillation with one constant restoring force": I'm not sure what this means. Perhaps you mean "a single restoring force"? For a simple pendulum, F = -kx (Hooke's Law), so F is not constant (its magnitude and direction change). And "typical" is an odd choice in this context: do you mean in comparison to other atmospheric oscillations? It might be simpler to just say that the QBO period is variable, and then go one to explain (as you do from line 6) what might be the causes of the variable period. Changed to "Furthermore, it is not a classic harmonic oscillation with one single restoring force, which leads to a variety of periods."

8: "these different aspects" –> "the different aspects of the QBO" Done.

8: "Figure" –> "e.g., Figure" Done.

9: suggest delete ", for example," Done.

10: add comma after "extent". Done.

11-17: On p. 2 you say, "The aims of this paper are to establish a set of standard metrics that comprehensively characterise the QBO." To be used by subsequent studies, the procedure for calculating these metrics needs to be unambiguous. I suggest you provide here a very clear algorithm (set of steps) that you used to calculate the metrics. Something like Charlton and Polvani 2007 ("A new look at SSWs, Part I"), Sec. 2b, is ideal: a numbered list of clearly described steps. Otherwise the reader has to fish through the text for the details, and it is easy for you to inadvertently omit some details. For example, in the caption of Fig 2 you say, "The Fourier harmonics around 2 years are averaged". You need to define the exact range of periods used. They are indicated by vertical lines in the left middle panel of Fig 2, but numbers need to be given so that the diagnostic is reproducible by others. It would also be worth mentioning that this introduces a dependence on the QBO period into all subsequent metrics that are based on the averaged Fourier amplitude, depending on the degree to which a given model's QBO period (which is variable) falls within the chosen range.

The definition of the metrics was clarified (p. 3, l. 5 - p. 4, l. 13); this particular comment is addressed as "The inverse of the minimum/maximum period is taken the upper/lower limit of the QBO Fourier harmonics." (p. 3, l. 23-26)

13: "height" –> "altitude". Similarly in Fig. 2 title of bottom panel. Done.

14: "QBO period" –> "distribution of QBO periods"

16-17: What is the min/max amplitude "from each QBO cycle"? Is it just the min/max wind, or wind shear? If so then remove "amplitude", or otherwise define how amplitude is calculated for a single QBO phase. Also state explicitly whether it's a wind amplitude,

or vertical wind shear amplitude, or both, that you're calculating. You say "shear zone", but you're discussing a time series of the wind at a single altitude.
Description changed to "The amplitude of the easterly/westerly phase in one QBO cycle are defined from the timeseries as the minimum/maximum wind value of a cycle. The values of each cycle are averaged to give the easterly/westerly amplitude". (p. 3, l. 21-11)
18: I think you mean the sum of squared amplitudes of Fourier harmonics that fall between the min and max QBO periods? State how the min/max QBO periods are determined: are these assumed values? (see comment for lines 11-17, above). This is potentially misleading because in the previous paragraph you said that the min/max QBO period is determined from the timeseries of u at h_max. But I assume you can't be referring to these periods here because h_max hasn't yet been defined, since you're describing here how you determine the latitude-height structure. So the order of presentation between the previous paragraph and this one is confusing. A clear, algorithmic description of how the metrics are calculated could fix this.
The definition of the metrics was clarified (p. 3, l. 5 - p. 4, l. 13); this particular comment is addressed as "The inverse of the minimum/maximum period is taken the upper/lower limit of the QBO Fourier harmonics." (p. 3, l. 23-26)
22: "maximum amplitude" –> "maximum" Not applicable anymore due to text changes.
23: Why is a fitted Gaussian used? Why not just use the latitude-altitude structure itself, as was done for the vertical depth? If a Gaussian is required for some reason (the reason should be stated), is it always a good fit? Does the fit quality vary amongst models? I'm worried that in comparing the values of this metric for different models, if a Gaussian is a good fit for one model but not another then the comparison may be less meaningful.
The Gaussian is a good fit for all models and is commonly used in QBO characterisation (e.g. Pascoe et al. (2005)).
23-24: "The QBO Fourier amplitude...": this sentence seems out of place here, since you have already referred to the maximum. Also, still unclear what is "maximum amplitude": is it just the maximum? The term "amplitude", here and leading up to this point, seems to be used carelessly. Amplitude is itself a metric, which can be defined in various ways, e.g. RMS amplitude of a time series. Not applicable anymore due to text changes.
27: "subsequent u values of opposite sign" –> "values of u having opposite sign at adjacent gridpoints" (or similar. "subsequent" seems the wrong word here) Done.
Page 4
6: "The progress... is noticeable": Do you mean from older to newer models in your set of models? If so, you could refer to Table 1 as indicating the vintages of the different models (by the year of the references given). Or, if you mean with respect to earlier results in the literature, please provide some specific comparisons.
Agreed that this was a vague statement. Changed to "The success of QBO simulation in GCMs is noticeable." (p. 5, l. 6)
11: insert "on average" after "QBO structure" Done.
11-12: Table 5 shows that the models and reanalysis disagree on h_max, i.e. the model error bars do not overlap the reanalysis value. So it seems incorrect to say that h_max in the models is realistic. This is also clear from Table 3, first column (h_max is 10 hPa for all but three models). The disagreement is consistent with your general result that the QBO in the models is shifted upward with respect to reanalyses. Yes, agreed. Deleted the statement about h_max.
16: insert " (Figure 4)" after "temperature amplitude" Done.
Page 5
2-3: Does the timing of phase transitions agree better between obs and reanalyes if you exclude some of the older reanalyses, such as NCEP1/2 and perhaps also JRA-25?
This was done. Even the more recent reanalyses have problems with phase transition representation

(see Kawatani et al (2016)).
5: In Table 5 I count ten models and eight reanalyses. Also, you assessed the observations (FUB winds). Changed.
8: "was established" and "was assessed" (previous paragraph used past tense - be consistent) Done.
11: I'm not sure where you commented on the variability of the QBO period in the models. Table 3 shows the min/max period, but plots of the distribution of periods would be more informative. Thanks for the suggestion. The plots have been included in the supplementary material.
12: "narrows" –> "is narrower", and "stronger than" –> "than in" Done.
14: I'm not sure this the correct way to state Haynes (1998)'s result. That paper shows that the QBO width not set by the width of the forcing when the imposed wave forcing is prescribed to have a very wide latitudinal distribution, designed not to impose a latitudinal scale on the QBO. I don't see that it rules out the actual forcing having a latitudinal distribution that might affect the QBO width. Changed sentence to " Even at the height of the maximum QBO, the modelled QBOs are too narrow, which suggests that there are shortcomings in modelling the factors, which influence the width as identified by Haynes (1998), such as the depth scale and the radiative damping."
You note that the width of the ITCZ and/or imposed gravity wave sources may play a role, and I agree.
22: "coupled" –> "coupled to" Done.
26-27: In Table 5, the standard deviation of descent rates for the models is the same for westerlies and easterlies. Either this statement is wrong or Table 5 is wrong. Statement removed.
31: If you mean that increased resolution leads to better representation of the wave forcing, perhaps change "(subsequently)" to "concomitantly" Done.
Table 1
- According to the text (p. 2), there are four CMIP5 models, not three. I believe CMCCCMS is also a CMIP5 model, and shares many similarities with MPI-ESM-MR. Please correct the caption. Done; thanks for spotting.
Table 3
- are confidence intervals for some of these columns appropriate? e.g.mean period. Done.
- why are the descent rates reported with fewer significant figures than the other metrics? Changed.
Table 4
- for temperature, lowest level (as in Table 3 for wind) would be a useful metric. Agreed. Included now.
Table 5
- For the reanalysis column, a number of the error values are zero.
- "Values are means and standard deviations of the metrics in Tables 3 and 4" –> "The mean and +/- one standard deviation of the metrics in Tables 3 and 4 are shown." Done.
- "excluding CMCC-CMS and both" –> "excluding both CMCC-CMS and" Both refers to the two UMUKCA models (-UCAM and -METO), so the word order should be correct.
- change "Depth" in the table to "Lowest level", to be consistent with Table 3 Done.
- why are the min/max periods not included? (all other metrics from Tables 3,4 are included) Included now.
Fig 1
- It would be helpful to expand this figure in the vertical (pressure) direction. Right now all the panels look kind of squished. Changed the panel formats.
- Label the middle panel to indicate that h_max is the blue horizontal line. Done.
- The blue and red lines in the middle panel are helpful. It's good how they correspond to the colours of the lines in the top, right, and bottom panels. But the dashed line style makes it easy to miss the colours. Perhaps make these solid lines. Done.
- It would help to add arrows between the panels indicating the algorithm for calculating

the metrics. That is, an arrow from the left (Fourier spectrum) pointing at the middle panel (latitude-altitude QBO amplitude), and then arrows from the middle panel point outward at the other three panels.
This is a reasonable suggestion. However, the authors feel that this would make the figure even busier and therefore kept the old format.
Fig 4
- Since the filled contours show the model bias (with respect to reanalyses), it would be more conventional to show the model-minus-reanalysis difference.
Plot/Description changed accordingly.
Fig 6
- This is subjective, but I find it very hard to compare the shapes of the three datasets in this format of plot. You might consider using a six-panel plot to show these results. You could have the phase transition direction as the row and the datasets as the columns (the plots could be narrower with only one dataset shown on each one).
The authors agreed on keeping the current presentation, which admittedly is dense, but therefore needs less space.

Thank you for your valuable comments.

Anonymous Referee #3

The QBO is the primary mode of variability in the Tropical stratosphere. The current paper aims to establish a set of standard metrics that comprehensively characterize the QBO. Subsequently the metrics are applied to 10 global circulation models, observations and reanalysis.
This paper is a very useful contribution, however I have some concerns and hence recommend major revisions.

Major Concerns:
1) The primary goal of this paper is to establish a standard set of metrics that can be used in the future. Ideally, the paper should include codes for calculating the metrics, so they are easily reproducible by other groups – hence point to a website from which such a diagnostic package can be downloaded. At the very least include very clear, step-by-step instructions on how the metrics were calculated should be included (without any ambiguity). The metrics presented here are reasonably well described, however there are lots of details in calculations, especially related to calculating the Fourier spectrum (step 1) which are omitted.
Definition of metrics was expanded/changed to an algorithmic description.(p. 3, l. 5 - p. 4, l. 13)

2) The paper somewhat lacks a description of what are the science goals motivating these metrics. The presented metrics seem useful to the general assessment of the representation of the QBO in global models, however they do not address aspects related to studying QBO related phenomena, such as QBO teleconnections for example. Hence, the use of these metrics is somewhat limited.
The purpose is to provide a phenomenological description of the QBO. They can be used in conjunction with teleconnection metrics (which, however, are well beyond the scope of this paper) to assess which QBO characteristics are more relevant for the interactions.
Added “These metrics were defined to be as simple as possible, yet meaningful in characterising the QBO morphologically. For robust and simple assessment of the QBO in models and observations, this study focusses on the large-scale morphology of the QBO rather than those (small-scale) dynamical processes involved in maintaining it.” (p. 2, l. 18-20)

3) The Fourier analysis is useful in certain respects for the assessment of the QBO (such as hmax, mean period), however from the mean and min/max QBO period values presented in Table 3 it is difficult to assess whether a model is getting the correct period distribution. The periods of the QBO vary between 20 and 35 months, and a simple histogram showing the number of times each period occurs would be more helpful in comparing observations to model output.
A corresponding figure was added to supplementary material.

4) It would be nice to see all the diagnostics for all the models in the appendix (ie.: Figure 2, Figure 4, Figure 6). The multi-model mean is nice to see and the numerical diagnostics are listed in Table 3, but the figures contain so much more information - it would be nice to see the complete set of metrics for all the models.
Figure 4 for models and period distribution added to supplementary material.

5) The metrics do not address the forcings of the QBO: gravity waves, resolved waves, vertical advection. It is possible for the QBO characteristics to be very close to observations, and for the forcing mechanisms to be unrealistic (ie.: lack of contribution from resolved waves, etc). Hence the addition of metrics addressing the momentum driving of the QBO would be a very important metric to add.
This would be beyond the scope of this paper. See comments to similar points above.

Minor Comments:
1) Page 2, Line 23: There is an inconsistency between ‘four CMIP5 models, and 5 CCMVAL models’ here and Table 1. In Table 1 only 3 models are listed as part of CMIP5: MIROC-ESM-CHEM, MPI-ESM-MR and HadGEM2-CC ; I believe the CMCCCMS should be included in the list of CMIP5 models in the caption of Table 1. Done.
2) Page 3, Line 3: ‘the period of the oscillation. . .’ - this should say ‘the mean period of the oscillation’ - it is well know that the period varies quite a bit as noted further in that paragraph Done.
3) Figure 2 caption: What is hmax ? Added “, where the equatorial QBO Fourier amplitude peaks,“ to the caption.
4) Page 5, Line 5: ‘eleven models’ – aren’t there only 10 in Table 4? True. Thanks for spotting.

Thank you for your feedback.

Defining metrics of the Quasi-Biennial Oscillation in global climate models

Verena Schenzinger[1], Scott Osprey[2,1], Lesley Gray[1], and Neal Butchart[3]

[1]Atmospheric, Oceanic and Planetary Physics, University of Oxford, UK
[2]National Centre for Atmospheric Science, UK
[3]Met Office Hadley Centre, Exeter, Devon, UK

Correspondence to: Verena Schenzinger (schenzinger@atm.ox.ac.uk)

Abstract. As the dominant mode of variability in the tropical stratosphere, the Quasi-Biennial Oscillation (QBO) has been subject to extensive research. Though there is a well developed theory of this phenomenon being forced by wave-mean flow interaction, simulating the QBO adequately in global climate models still remains difficult. This paper presents a set of metrics to characterise the QBO using a number of different reanalysis datasets and the FU Berlin radiosonde observation dataset. The same metrics are then calculated from Coupled Models Intercomparison Project 5 and Chemistry-Climate Model Validation Activity 2 simulations which included a representation of QBO-like behaviour to evaluate which aspects of the QBO are well captured by the models and which ones remain a challenge for future model development.

1 Introduction

After being referred to as a "mystery or freak" by one of its discoverers (Reed, 1967), the Quasi-Biennial Oscillation (QBO) now is accepted as the dominant pattern of variability in the equatorial stratosphere (Baldwin et al., 2001; Pascoe et al., 2005). Between 3 and 100 hPa, zonal wind at the equator is characterised by a pattern of descending easterly and westerly shear zones, with wind direction changing about every 14 months (see, for example, the ERA-Interim reanalysis and observations in Figure 1). The earliest regular observations of the equatorial stratosphere and hence the discovery of the QBO is credited to Ebdon (1960) and Reed et al. (1961). Angell and Korshover (1964), who named the phenomenon the "Quasi-Biennial Oscillation", pointed out oscillatory behaviour not only in zonal wind, but also in temperature, total ozone and tropopause height. The regularity of the oscillation makes it the most known repeatable mode of variability in the atmosphere, outside of the diurnal and seasonal cycles. Whether or not the QBO remains as regular in the present-day climate and under future climate change is an outstanding question (Osprey et al., 2016; Newman et al., 2016).

Early attempts to explain the driving mechanisms of the QBO failed in describing one or more of its main features, such as the quasi-biennial periodicity, the downward propagation or the roughly constant amplitude during the descent. Initial thoughts regarding the driving processes involved internal feedbacks, natural atmospheric modes, an unknown external process or a combination of those (Baldwin et al., 2001). The first study to explore possible forcing by gravity waves was by Lindzen and Holton (1968). They showed that vertically propagating waves could provide momentum for the QBO. This theory of wave-mean flow interaction was supported by a laboratory experiment, carried out by Plumb and McEwan (1978). They were able to produce a descending oscillation of the mean flow in a large annulus containing a salt-stratified fluid, the first practical demonstration of a laboratory analogue for the QBO. With the development of a theory of equatorial waves in the late 1960s, that was observationally confirmed (Maruyama, 1967; Wallace and Kousky, 1968), the work of Lindzen and Holton (1968) could be refined. Holton and Lindzen (1972) simulated a QBO-like oscillation in a simple one dimensional (1D) model, driven by vertically propagating Kelvin and Rossby-gravity waves that contribute westerly and easterly momentum forcing, respectively.

The first successful simulations of a reasonably realistic QBO were achieved in a 2D model by Gray and Pyle (1989) and in a 3D global climate model by Takahashi (1996). Follow-on studies describing simulations that captured a QBO were Horinouchi and Yoden (1998); Takahashi (1999); Scaife et al. (2000) and Hamilton et al. (2001). Adequate simulation of the QBO is affected by resolution (horizontal and vertical), parameterised gravity wave forcing from sub-grid scale waves (Giorgetta et al., 2006) and placement of the model lid (Lawrence, 2001; Osprey et al., 2013). However, there is not a simple model configuration that would guarantee a successful QBO simulation and despite there being a well established theory of the QBO, not all climate models can produce it. Of the 47 contributions submitted to the Coupled Model Intercomparison Project 5, CMIP5 (World Climate Research Programme, 2010), only five have a QBO-like signal (Lott et al., 2014)[1]. In the models submitted to the Chemistry-Climate Model Validation Activity (SPARC CCMVal, 2010) there are five out of fourteen, with three of them variants of the Met Office Unified Model (Butchart et al., 2011).

The aim of this paper is to establish a set of standard metrics that comprehensively characterise the QBO. These metrics were defined to be as simple as possible, yet meaningful in characterising the QBO morphologically. For robust and simple assessment of the QBO in models and observations, this study focusses on the large-scale morphology of the QBO rather than those (small-scale) dynamical processes involved in maintaining it. Using these characteristics, the performance of 10 historical model simulations is assessed and compared to observations and reanalysis datasets as the starting point of the World Climate Research Programme's (WCRP) Stratosphere-troposphere processes and their role in Climate (SPARC) QBO initiative (QBOi[2]) and SPARC Reanalysis Intercomparison Project (S-RIP). The purpose is to provide a benchmark for the current status of the representation of the QBO in global models against which new QBO-resolving simulations can be quantified.

2 Data

For this study, monthly means of zonally averaged zonal wind and temperature of four CMIP5 and five CCMVal-2 models as well as one from CMIP3 that internally produce a QBO were investigated. Table 1 lists these models and further details. Model data were obtained from the British Atmospheric Data Centre (BADC[3]). For comparison, the Berlin dataset (Freie Universität Berlin, 2015) of equatorial zonal wind from radiosonde observations covering 1956 to 2015 (Canton Island 1956-1967, Gan/Maledive Islands 1967-1975, Singapore 1967-2015) was analysed, as well as several reanalysis datasets (Table 2)

[1] Of these models, ten are resolving the stratosphere and include non-orographic gravity wave drag (Charlton-Perez et al., 2013), which are necessary ingredients for QBO simulation.

[2] http://users.ox.ac.uk/~astr0092/QBOi.html

[3] http://badc.nerc.ac.uk/home/index.html

made available through the SPARC Reanalysis Intercomparison Project (S-RIP) project[4]. When an average of more than one reanalysis was used, only the three relatively recent products (ERA Interim, MERRA, JRA55), comprising the years 1979-2009, were employed.

3 Definition of characteristic metrics

Figure 1 shows the equatorial zonal mean zonal wind for the different models, the ERA-Interim reanalysis and the FU Berlin dataset. In the models' stratosphere, QBO-like oscillations can be recognised. How much these resemble the observed QBO will be assessed based on a set of characteristic metrics. The most obvious one is the mean period; however the QBO has a structure in latitude and height and the behaviour of easterly and westerly phases differs. Furthermore, it is not a classic harmonic oscillation with one single restoring force, which leads to a variety of periods (Dunkerton, 2016). There might be an interaction with the semiannual oscillation or the 11 year solar cycle as well as the annual cycle in the troposphere that can influence timing of phase changes and descent of the shear zones. To assess the different aspects of the QBO that are seen in the zonal wind observations, we propose a set of characteristic metrics, including the height of the maximum amplitude, the latitudinal and vertical extent, and descent rates of each shear zone (Table 3, 1st row).

Figure 2 shows the process of metric derivation using the reanalyses mean (ERA-Interim, MERRA, JRA55) as an example. Derived values from the individual reanalyses, the FU Berlin dataset and model simulations are provided in Table 3. The metrics are defined as follows:

Paragraph rewritten completely

- The Fourier transformation of the equatorial zonal mean wind field (Figure 2, left panel) is calculated. The squares of the amplitudes between 26 and 30 months are added. The height of the maximum amplitude is taken as metric h_{max}.
- At h_{max}, the timeseries of $\bar{u}$ is used to find the QBO period, defined as the time between every other phase change (Figure 2, right panel)[5]. The minimum, maximum and mean of the periods are defined as QBO metrics. The months in which these phase changes occur are used to look for annual synchronisation of the QBO (Figure 6).
- The amplitude of the easterly/westerly phase in one QBO cycle are defined from the timeseries as the minimum/maximum wind value of a cycle. The values of each cycle are averaged to give the easterly/westerly amplitude.
- The inverse of the minimum/maximum period is taken the upper/lower limit of the QBO Fourier harmonics (Figure 2, left panel). The sum of the squares of the QBO amplitudes over the square root of the field variance gives the QBO Fourier amplitude. Doing this calculation for each grid point results in the QBO Fourier amplitude latitude-altitude structure (Figure 2, middle panel).

[4] http://s-rip.ees.hokudai.ac.jp/

[5] An alternative way to define a QBO period is presented by Wallace et al. (1993), who use the first two principle component timeseries of the stratospheric equatorial zonal wind in the approach. This has been applied to the FU Berlin dataset and results for the two methods are within each others error range: 28.0 ± 3.6 vs. 28.2 ± 4.4 months. For simplicity, the period metric has been defined from the raw zonal wind data.

- The vertical profile at the equator is calculated as the QBO Fourier amplitude for the zonal wind, averaged between 5° North and South (Figure 2, top panel). The vertical extent of the QBO is defined as the full depth at half maximum of the profile; the lowermost depth of the QBO (the lowermost level affected) is defined as the level of 10% of the maximum amplitude. Using the vertical profile, the value of the previously estimated h_{max} as the height of the maximum amplitude is validated.

- From the horizontal cross section at the height of the QBO maximum (Figure 2, bottom panel), the latitudinal extent (width) is defined by the full width at half maximum of a fitting Gaussian. The QBO Fourier amplitude is identified as the maximum amplitude, following Pascoe et al. (2005).

- The development of the profile of equatorial zonal wind serves to identify the descent rate of the shear zones. Figure 3 illustrates the procedure: at each point in time, the height of the sign change ($\bar{u} = 0$) of the wind profile is found by linear interpolation between two $\bar{u}$ values of opposite sign at adjacent gridpoints. The difference between the heights $\Delta h = h_{t+1} - h_t$, divided by the time resolution $\Delta t = 1$ month gives the descent rate at this timestep. The mean of the descent rates between 10 and 70hPa is calculated separately for the two shear zones as the mean over all values for a descending easterly/westerly.

The metrics for the temperature field are derived in an analogous way from the Fourier spectrum of the T timeseries. QBO temperature characteristics include the maximum Fourier amplitude, height of this maximum, depth of the QBO, latitudinal and vertical extent.

Section added

4 Error estimations

For metrics that are calculated as the mean over various cycles, the standard deviation ($\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})$) of the mean value is given as an error estimate. These are the mean period, the easterly/westerly wind amplitudes and the easterly/westerly descent rates.

The error of the minimum and maximum period is established following the method of surrogate timeseries presented by Christiansen (2010). First, the wind timeseries at h_{max} is subdivided into separate QBO cycles, with each cycle beginning at the minimum wind value between every other sign change of the wind. A long pseudo-QBO timeseries is constructed by concatenating 1000 randomly chosen cycles. From this timeseries, 100 samples of the same length as the original dataset are taken as surrogate QBO timeseries. The minimum and maximum period of these are estimated and the standard deviation is taken as the error estimate for the values. The error of the Fourier amplitude is calculated in the same way: First, the Fourier spectrum is calculated as in calculating the metric and the standard deviation of the 100 samples is used as error estimate for the Fourier amplitude.

The surrogate method does unfortunately not work where no clear QBO cycle can be defined - that is at levels below $\sim$ 70hPa or above $\sim$ 10hPa, or further away from the equator. Errors in metrics that are based on the Fourier amplitude outside the area dominated by the QBO (latitudinal and vertical extent, height of maximum, lowermost affected level) are mostly determined by the horizontal and vertical resolution of the model/reanalysis, which are given in Tables 1 and 2.

5 Model performance

Tables 3 and 4 list the characteristic metrics for all CMIP5 and CCMVal-2 models that have an internally generated QBO, for comparison with the reanalysis datasets and FUB observations (where possible). Table 5 compares the multi-model mean and the mean of the three most recent reanalyses. Figure 4 shows the multi-model and -reanalysis mean latitude-altitude QBO amplitude.

The success of QBO simulation in GCMs is noticeable: Most models represent the wind amplitude well compared to reanalyses and observations for both easterly and westerly QBO phases. Apart from 3 models (CMCC-CMS, UMUKCA-METO and UMUKCA-UCAM), the range of QBO periods is realistic (Table 3), with the multi-model mean not being significantly different from observations and reanalysis mean (Table 5).

A common model bias is a QBO that peaks slightly too high and does not descend low enough as seen in Figure 4. This indicates that the whole QBO structure on average is shifted slightly upwards. Even at the height of the maximum QBO amplitude, the simulated QBOs are too narrow in their latitudinal extent (Table 5). The reanalyses that resolve the atmosphere up to at least 1hPa (all except NCEP1/NCEP2) consistently show the maximum QBO at 20hPa, which is broadly in agreement with the FUB observations, given that the 15hPa level is not included in the reanalyses.

In the temperature field, half of the models peak at a realistic height (20-30hPa), whereas the other half peaks too high (~5 hPa) which leads on average to an elongated structure in height for the QBO temperature amplitude (Figure 4). Again, the difference between the model and the reanalysis mean shows a shift of the QBO structure upwards. Additionally, there is a slight overestimation of the QBO temperature amplitude at subtropical latitudes (15°-30°) in the models. Exclusion of models with obvious shortcomings in QBO modelling as seen by unrealistic periods does not significantly improve these biases (Table 5).

There is a slight asymmetry in the descent rates of easterly and westerly shear zones in models, but it is not as pronounced as in the observations/reanalyses, where the westerlies descend about twice as fast as the easterlies. Figure 5 shows the easterly and westerly descent rates for each model and reanalysis dataset as well as the mutli-model/reanalysis mean and standard deviations. Even the model with the fastest descending westerlies still has a slower descent rate than the observations and the slowest reanalysis dataset. Most of the models have comparable westerly and easterly descent rates, with UMSLIMCAT even reversing the asymmetry towards faster easterlies. While within reanalyses and the FUB observations, the standard deviation in the easterly descent rate is usually slightly larger than in the westerly descent rate, the inter-model/-reanalysis discrepancy is higher for descending westerlies. Models show similar standard deviations for both westerly and easterly descent rate, which can also be seen in a more uniform descent of both shear zones and less prominent stalling features compared to the observations (Figure 1).

Figure 6 shows the timing of the phase change at the height of the maximum QBO amplitude. For both west to east and east to west transitions, there is a seasonal modulation in the models with more changes occuring in boreal spring and autumn, but this modulation is not as prominent as in the FUB observations, where west to east transitions are favoured in May and November and east to west transitions are slightly more common in November. Reanalyses favour west to east transitions in October and east to west transitions in December. However, with only 29 FUB observational cycles and 39 (3x13) in total in the reanalyses to compare, no conclusive statement about the significance of the difference between models and reanalyses/observations can be made. It is, however, intriguing that the distributions of the west-east and east-west transitions look similar in the models, but not in the observations/reanalyses.

6 Discussion and Conclusion

The representation of the stratospheric zonal mean wind and temperature fields in ten models, eight reanalysis datasets and the FU Berlin observations (wind only) was assessed in this paper. It is a positive development that an increasing number of global climate models resolve the stratosphere well enough to show an oscillation in zonal mean zonal wind that resembles the observed QBO.

A set of metrics to characterise the quality of these simulations was established and the model performance was evaluated using reanalyses and the FUB observational radiosonde dataset as reference. Some typical features of the QBO are well represented, such as the asymmetry in easterly/westerly amplitude, the latitudinal confinement around the equator and the vertical extent. Apart from three models, the mean period and its variability is captured well. However, the QBO in all models is shifted upwards in height compared to reanalyses and narrows in latitude in the lower stratosphere stronger than the reanalyses (Figure 4). Even at the height of the maximum QBO, the modelled QBOs are too narrow, which suggests that there are shortcomings in modelling the factors, which influence the width as identified by Haynes (1998), such as the depth scale and the radiative damping. The parametrization of the gravity wave sources or the width of the inter-tropical convergence zone might play a role as well. However, the disagreement between reanalyses is also greatest at low latitudes as noted by Kawatani et al. (2016), a finding they explain by the small equatorial Coriolis parameter and sparse observations.

The discrepancy between the timing of phase transitions in the reanalyses and observations (Figure 6) was also pointed out by Kawatani et al. (2016). Model behaviour differs even more from the observations, with similar phase transition distribution for both east-west and west-east transitions. Kawatani et al. (2016) suggest that weak forcing by resolved waves contributes to the bias in reanalysis, a mechanism that might also lead to the discrepancy in models. Furthermore, parametrized gravity waves in the models used in this study are not coupled to the main generation processes in the atmosphere, such as tropical convection, which might explain why the annual variation in phase transitions is not as prominent as in the observations.

Insufficient wave forcing might also be responsible for the lack of difference between easterly and westerly descent rates. In observations, westerlies descend on average about twice as fast as easterlies, whereas in models the difference in rates is not significant, with the westerlies descending too slowly. The standard deviation of the multi-reanalysis mean is higher for westerly than for easterly descent rates, a result that also points towards disagreement in the underlying westerly forcing.

In summary, there has been substantial improvement in simulating the tropical stratosphere in global climate models, with

QBO-like oscillations being represented in a growing number of models. The characteristic metrics defined here present the possibility of quickly assessing the quality of a simulation. With improving model resolution and (concomitantly) the representation of wave forcing, GCMs are very likely to simulate a more realistic QBO.

7 Data availability

CMIP5 and CCMVal-2 climate model data was downloaded from the British Atmospheric Data Centre (BADC), HadGEM1 data can be obtained from SMO. For reanalysis data please contact Masatomo Fujiwara, who prepared it for the SPARC reanalysis intercomparison project (S-RIP), or the respective centre as listed here http://s-rip.ees.hokudai.ac.jp/resources/links.html.

Acknowledgements. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups for producing and making available their model output. We thank the modelling groups of the SPARC Chemistry-Climate Model Validation (CCMVal-2) for producing and making available their model output. We thank the British Atmospheric Data Centre (BADC) for providing access to CMIP5 and CCMVal-2 data and the SPARC reanalysis intercomparison project (S-RIP) for the reanalysis datasets.

SMO and LJG would like to to acknowledge funding from the National Environment Research Council projects QBOnet (NE/M005828/1) and GOTHAM (NE/P006779/1), NB was supported by the Joint UK BEIS/DefraMet Office Hadley Centre Climate Programme (GA01101). The authors would also like to thank Adam Scaife, Andrew Bushell, Jeff Knight and Martin Andrews for their valuable contributions to early discussions helping to motivate this study. We thank James Anstey, Kevin Hamilton, Thomas Krismer, John McCormack, Marv Geller and Tiehan Zou for early discussions on and contributions to the SPARC-QBOi questionnaire.

References

Angell, J. K. and Korshover, J.: Quasi-biennial variations in temperature, total ozone and tropopause height, Journal of the Atmospheric Sciences, 21, 479–492, 1964.

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, Reviews of Geophysics, 39, 179–229, doi:10.1029/1999RG000073, 2001.

Bushell, A. C., Jackson, D. R., Butchart, N., Hardiman, S. C., Hinton, T. J., Osprey, S. M., and Gray, L. J.: Sensitivity of GCM tropical middle atmosphere variability and climate to ozone and parameterized gravity wave changes, Journal of Geophysical Research: Atmospheres, 115, D15 101, 2010.

Butchart, N., Charlton-Perez, A. J., Cionni, I., Hardiman, S. C., Haynes, P. H., Krüger, K., Kushner, P. J., Newman, P. A., Osprey, S. M., Perlwitz, J., Sigmond, M., Wang, L., Akiyoshi, H., Austin, J., Bekki, S., Baumgaertner, A., Braesicke, P., Brühl, C., Chipperfield, M., Dameris, M., Dhomse, S., Eyring, V., Garcia, R., Garny, H., Jöckel, P., Lamarque, J.-F., Marchand, M., Michou, M., Morgenstern, O., Nakamura, T., Pawson, S., Plummer, D., Pyle, J., Rozanov, E., Scinocca, J., Shepherd, T. G., Shibata, K., Smale, D., Teyssèdre, H., Tian, W., Waugh, D., and Yamashita, Y.: Multimodel climate and variability of the stratosphere, Journal of Geophysical Research: Atmospheres, 116, D05 102, 2011.

Charlton-Perez, A. J., Baldwin, M. P., Birner, T., Black, R. X., Butler, A. H., Calvo, N., Davis, N. A., Gerber, E. P., Gillett, N., Hardiman, S., Kim, J., Krüger, K., Lee, Y.-Y., Manzini, E., McDaniel, B. A., Polvani, L., Reichler, T., Shaw, T. A., Sigmond, M., Son, S.-W., Toohey, M., Wilcox, L., Yoden, S., Christiansen, B., Lott, F., Shindell, D., Yukimoto, S., and Watanabe, S.: On the lack of stratospheric dynamical variability in low-top versions of the CMIP5 models, Journal of Geophysical Research: Atmospheres, 118, 2494–2505, 2013.

Christiansen, B.: Stratospheric Bimodality: Can the Equatorial QBO Explain the Regime Behaviour of the NH Winter Vortex?, Journal of Climate, 23, 3953–3966, 2010.

Dunkerton, T. J.: The Quasi-Biennial Oscillation of 2015-16: Hiccup or Death Spiral?, Geophys. Res. Lett., 43, 10,547–10,552, 2016.

Ebdon, R. A.: Notes on the wind flow at 50mb in tropical and subtropical regions in January 1957 and in 1958, Quarterly Journal of the Royal Meteorological Society, 86, 540–542, 1960.

Ebita, A., Kobayashi, S., Ota, Y., Moriya, M., Kumabe, R., Onogi, K., Harada, Y., Yasui, S., Miyaoka, K., Takahashi, K., Kamahori, H., Kobayashi, C., Endo, H., Soma, M., Oikawa, Y., and Ishimizu, T.: The Japanese 55-year Reanalysis "JRA-55": an interim report, SOLA, 7, 149–152, this article is an interim report of JRA-55 as of 2011. A comprehensive report of JRA-55 is under preparation to be submitted to J. Meteor. Soc. Japan (as of November 2013), 2001.

Freie Universität Berlin: The Quasi-Biennial-Oscillation (QBO) Data Serie, http://www.geo.fu-berlin.de/en/met/ag/strat/produkte/qbo/, accessed: 2016-07-01, 2015.

Giorgetta, M. A., Manzini, E., E., R., Esch, M., and Bengtsson, L.: Climatology and Forcing of the Quasi-Biennial Oscillation in the MAECHAM5 Model, Journal of Climate, 19, 3882–3901, 2006.

Gray, L. J. and Pyle, J. A.: A two-dimensional model of the quasi-biennial oscillation in ozone, Journal of the Atmospheric Sciences, 56, 203–220, 1989.

Hamilton, K., Wilson, R. J., and Hemler, R. S.: Spontaneous Stratospheric QBO-like Oscillations Simulated by the GFDL SKYHI General Circulation Model, Journal of the Atmospheric Sciences, 58, 3271–3292, 2001.

Hardiman, S. C., Butchart, N., Osprey, S. M., Gray, L. J., Bushell, A. C., and Hinton, T. J.: The Climatology of the Middle Atmosphere in a Vertically Extended Version of the Met Office's Climate Model. Part I: Mean State, Journal of the Atmospheric Sciences, 67, 1509–1525, 2010.

Hardiman, S. C., Butchart, N., Hinton, T. J., Osprey, S. M., and Gray, L. J.: The Effect of a Well-Resolved Stratosphere on Surface Climate: Differences between CMIP5 Simulations with High and Low Top Versions of the Met Office Climate Model, Journal of Climate, 25, 7083–7099, 2012.

Haynes, P. H.: The latitudinal structure of the quasi-biennial oscillation, Quaterly Journal of the Royal Meteorological Society, 124, 2645–2670, 1998.

Hines, C. O.: Doppler-spread parameterization of gravity wave momentum deposition in the middle atmosphere. Part 1: Basic formulation, Journal of Atmospheric and Solar-Terrestrial Physics, 59, 371–386, 1997a.

Hines, C. O.: Doppler-spread parameterization of gravity wave momentum deposition in the middle atmosphere. Part 2: Broad and quasi-monochromatic spectra, and implementation, Journal of Atmospheric and Solar-Terrestrial Physics, 59, 387–400, 1997b.

Holton, J. R. and Lindzen, R. S.: An updated theory for the quasi-biennial cycle of the tropical stratosphere, Journal of the Atmospheric Sciences, 29, 1076–1080, 1972.

Horinouchi, T. and Yoden, S.: Wave–Mean Flow Interaction Associated with a QBO-like Oscillation Simulated in a Simplified GCM, Journal of the Atmospheric Sciences, 55, 502–526, 1998.

Jöckel, P., Tost, H., Pozzer, A., Bruehl, C., Buchholz, J., Ganzeveld, L., Hoor, P., Kerkweg, A., Lawrence, M., Sander, R., Steil, B., Stiller, G., Tanarhte, M., Taraborrelli, D., van Aardenne, J., and Lelieveld, J.: The atmospheric chemistry general circulation model ECHAM5/MESSy1:Consistent simulation of ozone from the surface to the mesosphere, Atmospheric Chemistry and Physics, 6, 5067–5104, 2006.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., and Reynolds, R.: The NCEP/NCAR 40-Year Reanalysis Project, Bulletin of the American Meteorological Society, 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), Bulletin of the American Meteorological Society, 83, 1631–1643, 2002.

Kawatani, Y., Hamilton, K., Miyazaki, K., Fujiwara, M., and Anstey, J. A.: Representation of the tropical stratospheric zonal wind in global atmospheric reanalyses, Atmospheric Chemistry and Physics, 16, 6681–6699, 2016.

Kistler, R., Collins, W., Saha, S., White, G., Woollen, J., Kalnay, E., Chelliah, M., Ebisuzaki, W., Kanamitsu, M., Kousky, V., van den Dool, H., Jenne, R., and Fiorino, M.: The NCEP–NCAR 50–Year Reanalysis: Monthly Means CD–ROM and Documentation, Bulletin of the American Meteorological Society, 82, 247–267, 2001.

Krismer, T. R. and Giorgetta, M. A.: Wave Forcing of the Quasi-Biennial Oscillation in the Max Planck Institute Earth System Model, Journal of the Atmospheric Sciences, 71, 1985–2006, 2014.

Lawrence, B. N.: A gravity-wave induced quasi-biennial oscillation in a three-dimensional mechanistic model, Quarterly Journal of the Royal Meteorological Society, 127, 2005–2021, 2001.

Lindzen, R. S. and Holton, J. R.: A Theory of the quasi-biennial oscillation, Journal of the Atmospheric Sciences, 25, 1095–1107, 1968.

Lott, F., Denvil, S., Butchart, N., Cagnazzo, C., Giorgetta, M. A., Hardiman, S. C., Manzini, E., Krismer, T., Duvel, J.-P., Maury, P., Scinocca, J. F., Watanabe, S., and Yukimoto, S.: Kelvin and Rossby-gravity wave packets in the lower stratosphere of some high-top CMIP5 models, Journal of Geophysical Research: Atmospheres, 119, 2156–2173, 2014.

Manzini, E., Giorgetta, M. A., Esch, M., Kornblueh, L., and Roeckner, E.: The Influence of Sea Surface Temperatures on the Northern Winter Stratosphere: Ensemble Simulations with the MAECHAM5 Model, Journal of Climate, 19, 3863–3881, 2006.

Maruyama, T.: Large-scale disturbances in the equatorial lower stratosphere, Journal of the Meteorological Society of Japan, 45, 169–199, 1967.

Morgenstern, O., Braesicke, P., O'Connor, F. M., Bushell, A. C., Johnson, C. E., Osprey, S. M., and Pyle, J. A.: Evaluation of the new UKCA climate-composition model – Part 1: The stratosphere, Geoscientific Model Development Discussions, 2, 43–57, 2009.

Newman, P. A., Coy, L., Pawson, S., and Lait, L. R.: The anomalous change in the QBO in 2015-2016, Geophysical Research Letters, 43, 8791–8797, 2016.

Onogi, K., 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., and Taira, R.: The JRA-25 Reanalysis, Journal of the Meteorological Society of Japan, 85, 369–432, 2007.

Osprey, S. M., Gray, L. J., Hardiman, S. C., Butchart, N., Bushell, A. C., and Hinton, T. J.: The Climatology of the Middle Atmosphere in a Vertically Extended Version of the Met Office's Climate Model. Part II: Variability, Journal of the Atmospheric Sciences, 67, 3637–3651, 2010.

Osprey, S. M., Gray, L. J., Hardiman, S. C., Butchart, N., and Hinton, T. J.: Stratospheric Variability in Twentieth-Century CMIP5 Simulations of the Met Office Climate Model: High Top versus Low Top, Journal of Climate, 26, 1595–1606, 2013.

Osprey, S. M., Butchart, N., Knight, J. R., Scaife, A. A., Hamilton, K., Anstey, J. A., Schenzinger, V., and Zhang, C.: An unexpected disruption of the atmospheric quasi-biennial oscillation, Science, 2016.

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, Journal of Geophysical Research: Atmospheres, 110, doi:10.1029/2004JD004941, 2005.

Plumb, R. A. and McEwan, A. D.: The Instability of a Forced Standing Wave in a Viscous Stratified Fluid: A Laboratory Analogue of the Quasi-Biennial Oscillation, Journal of the Atmospheric Sciences, 35, 1827–1839, 1978.

Reed, R. J.: The structure and dynamics of the 26-month oscillation in tropical latitudes, in: Proceedings of the International Symposium on the Dynamics of Large-Scale Processes in the Atmosphere, pp. 376–387, Russian Academy of Science, Moscow, 1967.

Reed, R. J., Campbell, W. J., Rasmussen, L. A., and Rogers, D. G.: Evidence of a downward-propagating, annual wind reversal in the equatorial stratosphere, Journal of Geophysical Research, 66, 813–818, 1961.

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, Journal of Climate, 24, 3624–3648, 2011.

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., H.-Ming, Juang, H., Sela, J., Iredell, M., Treadon, R., Kleist, D., van Delst, P., Keyser, D., Derber, J., Ek, M., Meng, J., Wei, H., Yang, R., Lord, S., van den Dool, H., 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, Bulletin of the American Meteorological Society, 91, 1015–1057, 2010.

Scaife, A. A., Butchart, N., Warner, C. D., Stainforth, D., Norton, W., and Austin, J.: Realistic quasi-biennial oscillations in a simulation of the global climate, Geophysical Research Letters, 27, 3481–3484, 2000.

Schmidt, H., Rast, S., Bunzel, F., Esch, M., Giorgetta, M., Kinne, S., Krismer, T., Stenchikov, G., Timmreck, C., Tomassini, L., and Walz, M.: Response of the middle atmosphere to anthropogenic and natural forcings in the CMIP5 simulations with the Max Planck Institute Earth system model, Journal of Advances in Modeling Earth Systems, 5, 98–116, 2013.

Shibata, K. M. and Deushi, M.: Long-term variations and trends in the simulation of the middle atmosphere 1980-2004 by the chemistry-climate model of the Meteorological Research Institute, Annals of Geophysics, 26, 1299–1326, 2008a.

Shibata, K. M. and Deushi, M.: Simulation of the stratospheric circulation and ozone during the recent past (1980-2004) with the MRI chemistry-climate model, in: CGER's Supercomputer Monograph Report, vol. 13, p. 154 pp., NIES, Japan, 2008b.

SPARC CCMVal: SPARC Report on the Evaluation of Chemistry-Climate Models, sPARC Report No. 5, WCRP-132, WMO/TD-No. 1526, 2010.

Takahashi, M.: Simulation of the stratospheric quasi-biennial oscillation using a general circulation model , Geophysical Research Letters, 23, 661–664, 1996.

Takahashi, M.: Simulation of the Quasi-Biennial Oscillation in a general circulation model, Geophysical Research Letters, 26, 1307–1310, 1999.

Tian, W. and Chipperfield, M. P.: A new coupled-chemistry model for the stratosphere: The importance of coupling for future O3-climate predictions, Quarterly Journal of the Royal Meteorological Society, 131, 281–303, 2005.

Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D. C., 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., Berg, L. V. D., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, 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., Ster, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J.: The ERA-40 re-analysis, Quarterly Journal of the Royal Meteorological Society, 131, 2961–3012, 2005.

Wallace, J. M. and Kousky, V. E.: Observational evidence of Kelvin waves in the tropical stratosphere, Journal of the Atmospheric Sciences, 25, 900–907, 1968.

Wallace, J. M., Panetta, R. L., and Estberg, J.: Representation of the Equatorial Stratospheric Quasi-Biennial Oscillation in EOF Phase Space, Journal of the Atmospheric Sciences, 50, 1751–1762, 1993.

Warner, C. D. and McIntyre, M. E.: An ultrasimple spectral parametrization for nonorographic gravity waves, Journal of the Atmospheric Sciences, 85, 1837–1857, 2001.

Watanabe, S. and Kawatani, Y.: Sensitivity of the QBO to Mean Tropical Upwelling under a Changing Climate Simulated with an Earth System Model, Journal of the Meteorological Society of Japan. Ser. II, 90A, 351–360, 2012.

Watanabe, S., Hajima, T., Sudo, K., Nagashima, T., Takemura, T., Okajima, H., Nozawa, T., Kawase, H., Abe, M., Yokohata, T., Ise, T., Sato, H., Kato, E., Takata, K., Emori, S., and Kawamiya, M.: MIROC-ESM 2010: model description and basic results of CMIP5-20c3m experiments, Geoscientific Model Development, 4, 845–872, 2011.

World Climate Research Programme: WCRP Coupled Model Intercomparison Project - Phase 5, Special Issue of the CLIVAR Exchanges Newsletter, 15, 2010.

Model	Reference	Resolution	GW scheme	Length
HadGEM1	Osprey et al. (2010)	N96 L60	W &M	50 years
	Hardiman et al. (2010)			
	Bushell et al. (2010)			
MIROC-ESM-CHEM	Watanabe et al. (2011),	T42 L68	Hines	156 years
	Watanabe and Kawatani (2012)			
MPI-ESM-MR	Schmidt et al. (2013),	T63 L95	Hines	156 years
	Krismer and Giorgetta (2014)			
HadGEM2-CC	Osprey et al. (2013)	1.25° x 1.875° L60	W &M	374 years
	Hardiman et al. (2012)			
CMCC-CMS	Manzini et al. (2006),	T63 L95	Hines	156 years
	Giorgetta et al. (2006)			
EMAC	Jöckel et al. (2006)	T42 L90	Hines	41 years
MRI	Shibata and Deushi (2008a),	T42 L68	Hines	47 years
	Shibata and Deushi (2008b)			
UMSLIMCAT	Tian and Chipperfield (2005)	2.50° x 3.75° L64	W &M	55 years
UMUKCA-METO	Morgenstern et al. (2009)	2.50° x 3.75° CP60	W &M	47 years
UMUKCA-UCAM	Morgenstern et al. (2009)	2.50° x 3.75° CP60	W &M	45 years

Table 1. Climate models used in the study. HadGEM1 was part of CMIP3, MIROC-ESM-CHEM, MPI-ESM-MR, HadGEM2-CC and CMCC-CMS were part of CMIP5, the rest are CCMVal-2 models. CMIP5 models are runs with a coupled ocean, HadGEM1 and the CCMVal-2 models are atmosphere only runs. The gravity wave (GW) parametrisation schemes are based on either the Ultra-Simple Spectral Parametrisation Warner and McIntyre (2001) (W & M) or the Doppler Spread Parametrisation scheme Hines (1997a, b) (Hines).

Reanalysis	Reference	Resolution of forecast model
ERA40	Uppala et al. (2005)	T_L159 and N80 reduced Gaussian, L60
ERA Interim	Uppala et al. (2005)	T_L255 and N128 reduced Gaussian, L60
MERRA	Rienecker et al. (2011)	0.66° lon x 0.5° lat; 72 sigma levels
JRA25	Onogi et al. (2007)	T106 L40
JRA55	Ebita et al. (2001)	T_L319 L60
CFSR	Saha et al. (2010)	T382 L64
NCEP1	Kalnay et al. (1996)	T62 L28
	Kistler et al. (2001)	
NCEP2	Kanamitsu et al. (2002)	T62 L28

Table 2. Reanalysis datasets used in the study. The period is 1979-2009 for all reanalyses except ERA40, which covers 1958-2001.

Model/Reanalysis	Height of maximum (hPa) *	Fourier amplitude (m/s)	Latitudinal extent (°) *	Vertical extent (km) *	Lowest level (hPa)	Period (months) Min	Max	Mean	Amplitude (m/s) East	West	Descent rate (km/month) East	West
HadGEM1	10	14.7 ± 1.0	17.9	18.1	78	26.2 ± 3.1	41.3 ± 2.8	33.8 ± 4.4	-35.5 ± 1.4	16.0 ± 2.2	0.67 ± 0.50	0.72 ± 0.53
HadGEM2-CC	10	18.6 ± 0.9	18.7	16.7	83	21.1 ± 1.3	31.2 ± 1.4	26.5 ± 2.5	-34.8 ± 1.5	21.4 ± 2.3	0.76 ± 0.52	1.10 ± 0.50
MIROC-ESM-CHEM	15	16.0 ± 0.2	19.6	17.9	76	22.3 ± 0.8	31.8 ± 0.9	26.1 ± 2.0	-34.7 ± 2.1	14.4 ± 1.7	0.68 ± 0.39	0.85 ± 0.55
MPI-ESM-MR	10	22.1 ± 0.5	20.2	18.3	84	23.5 ± 1.7	57.0 ± 7.8	30.3 ± 4.8	-38.1 ± 1.9	26.1 ± 1.5	0.58 ± 0.49	0.90 ± 0.68
CMCC-CMS	10	14.4 ± 0.8	18.8	19.0	68	22.6 ± 2.2	153.5 ± 18.6	47.9 ± 29.9	-28.3 ± 7.6	26.4 ± 2.0	0.44 ± 0.55	0.59 ± 1.41
EMAC	15	13.2 ± 0.6	21.0	18.5	76	24.1 ± 1.8	35.0 ± 2.1	28.4 ± 3.3	-29.0 ± 2.3	14.3 ± 1.4	0.73 ± 0.63	0.96 ± 0.59
MRI	20	13.3 ± 1.0	20.6	21.5	81	25.3 ± 1.8	38.3 ± 2.3	28.3 ± 3.1	-27.4 ± 1.1	12.5 ± 1.3	0.57 ± 0.44	1.03 ± 0.88
UMSLIMCAT	10	15.2 ± 1.2	18.8	22.7	77	28.5 ± 1.9	35.1 ± 1.6	31.2 ± 1.8	-36.7 ± 2.4	23.0 ± 3.2	0.83 ± 0.58	0.57 ± 0.45
UMUKCA-METO	10	11.1 ± 2.1	18.2	15.6	80	53.2 ± 7.1	62.4 ± 4.4	56.8 ± 3.3	-34.0 ± 1.8	17.0 ± 1.5	0.25 ± 0.47	0.52 ± 0.53
UMUKCA-UCAM	10	12.9 ± 0.8	18.2	16.7	87	36.0 ± 6.1	66.3 ± 4.5	49.5 ± 9.5	-34.8 ± 1.4	17.8 ± 0.9	0.33 ± 0.46	0.56 ± 0.45
ERA 40	20	14.6 ± 0.6	21.7	17.3	89	19.6 ± 1.7	34.7 ± 2.0	28.3 ± 3.8	-35.6 ± 2.4	13.4 ± 2.3	0.74 ± 0.71	1.32 ± 0.72
ERA-Interim	20	14.6 ± 1.0	20.9	15.1	86	22.9 ± 1.7	34.7 ± 2.5	28.0 ± 3.5	-34.5 ± 2.3	16.0 ± 1.4	0.79 ± 0.78	1.28 ± 0.71
MERRA	20	15.1 ± 1.1	20.8	17.7	89	22.5 ± 1.5	34.8 ± 1.8	28.0 ± 3.7	-34.6 ± 2.5	16.4 ± 1.9	0.74 ± 0.69	1.16 ± 0.68
JRA25	20	14.1 ± 1.3	20.9	19.4	87	22.9 ± 2.1	35.2 ± 2.6	28.0 ± 3.6	-35.4 ± 2.1	16.4 ± 2.0	0.73 ± 0.68	1.56 ± 0.68
JRA55	20	14.5 ± 1.2	21.3	17.7	88	22.9 ± 1.6	35.6 ± 2.9	28.0 ± 3.7	-35.6 ± 2.3	16.1 ± 1.5	0.69 ± 0.67	1.44 ± 0.59
CFSR	20	14.0 ± 0.8	21.4	16.5	83	23.2 ± 1.6	34.8 ± 2.2	27.9 ± 3.5	-34.9 ± 2.3	14.7 ± 2.8	0.75 ± 0.83	1.18 ± 0.65
NCEP1	20	11.7 ± 1.1	20.1	-	80	22.9 ± 1.5	34.5 ± 2.8	27.9 ± 3.5	-27.9 ± 3.4	15.3 ± 1.4	0.82 ± 0.79	1.55 ± 0.68
NCEP2	10	12.1 ± 0.8	20.6	-	81	22.5 ± 1.6	35.1 ± 2.7	27.7 ± 4.2	-35.6 ± 3.8	10.0 ± 3.5	0.84 ± 0.84	1.68 ± 0.87
Observations	15	15.7 ± 0.6	-	-	-	20.4 ± 1.8	36.9 ± 2.3	28.2 ± 4.5	-32.8 ± 3.2	18.3 ± 2.8	0.66 ± 0.75	1.24 ± 0.65

Table 3. Calculated QBO metrics for all models and reanalyses. Where possible, the value from the observations is given as well. However, these only consist of one timeseries at the equator (hence no latitudinal information) and are available between 10 and 70hPa for the time 1956-2015, so the depth and vertical extent could not be assessed. The highest level NCEP1 and NCEP2 is 10hPa, which accounts for the missing value in vertical extent. Errors of mean period, easterly/westerly amplitudes and descent rates are standard errors based on averaging over QBO cycles. The error in the Fourier amplitude, min/max period is based on surrogate timeseries.

* The error of these parameters is determined by the grid spacing (refer to Tables 1 and 2).

Added error estimates to numbers

Model/Reanalysis	Height of maximum (hPa) *	Fourier amplitude (K)	Latitudinal extent (°) *	Vertical extent (km) *	Lowest Level (hPa) *
HadGEM1	15	0.7 ± 0.1	12.9	20.0	89
HadGEM2-CC	6	1.0 ± 0.1	14.4	19.2	96
MIROC-ESM-CHEM	7	1.4 ± 0.0	13.8	16.3	69
MPI-ESM-MR	5	1.7 ± 0.0	15.2	20.8	82
CMCC-CMS	5	1.1 ± 0.1	16.0	22.9	85
EMAC	20	1.2 ± 0.1	15.7	17.7	85
MRI	30	0.9 ± 0.1	15.2	19.6	97
UMSLIMCAT	20	1.0 ± 0.1	13.2	21.3	85
UMUKCA-METO	30	0.7 ± 0.1	12.8	19.6	113
UMUKCA-UCAM	30	0.8 ± 0.1	13.6	18.5	115
ERA 40	30	1.3 0.1	16.2	14.2	97
ERA-Interim	30	1.3 0.1	16.8	14.9	89
MERRA	30	1.3 0.1	16.8	14.8	88
JRA25	30	1.1 0.2	15.8	17.4	89
JRA55	30	1.3 0.1	16.9	13.7	88
CFSR	20	1.2 0.1	17.4	15.2	85
NCEP1	30	0.8 0.1	15.3	-	85
NCEP2	20	0.8 0.1	27.7	-	87

Table 4. Characteristic QBO metrics calculated from the zonal mean temperature. Values for models and reanalyses are listed; there is no comparable observational dataset.

* The error of these parameters is determined by the grid spacing (refer to Tables 1 and 2).

ZM Zonal Wind	Model mean	Model mean (ex)	Reanalysis mean
Height of maximum (hPa)	12.0 ± 3.5	13.3 ± 4.1	20.0 ± 0.0
Fourier amplitude (m/s)	15.1 ± 3.2	16.3 ± 3.5	14.8 ± 0.3
Latitudinal extent (°)	19.2 ± 1.1	19.7 ± 1.2	21.0 ± 0.3
Vertical extent (km)	18.5 ± 2.2	18.5 ± 1.6	16.8 ± 1.5
Lowest Level (hPa)	79.1 ± 5.2	79.8 ± 3.6	87.6 ± 1.7
Mean Period (months)	35.9 ± 11.2	28.9 ± 2.8	28.0 ± 0.0
Min Period (months)	28.3 ± 9.7	23.8 ± 1.9	22.7 ± 0.2
Max Period (months)	55.2 ± 36.9	39.1 ± 9.6	35.0 ± 0.5
Amplitude Easterly	-33.3 ± 3.7	-33.3 ± 4.1	-34.0 ± 1.9
Amplitude Westerly	18.9 ± 5.0	17.5 ± 5.2	15.8 ± 0.8
Descent rate Easterly	0.6 ± 0.2	0.7 ± 0.1	0.7 ± 0.0
Descent rate Westerly	0.8 ± 0.2	0.9 ± 0.1	1.3 ± 0.1
ZM Temperature			
Height of maximum (hPa)	16.8 ± 10.7	13.8 ± 9.9	30.0 ± 0.0
Fourier amplitude (m/s)	1.1 ± 0.3	1.2 ± 0.4	1.3 ± 0.0
Latitudinal extent (°)	14.3 ± 1.2	14.5 ± 1.0	16.9 ± 0.0
Vertical extent (km)	19.6 ± 1.9	18.9 ± 1.6	14.5 ± 0.7
Lowest Level (hPa)	91.5 ± 14.0	86.3 ± 10.2	88.0 ± 0.7

Table 5. Characteristic QBO metrics in reanalyses and models. The mean and ± one standard deviation of the metrics in Tables 3 and 3 are shown. The multi-model mean was calculated from all models (* excluding CMCC-CMS and both UMUKCA models for obvious shortcomings in QBO modelling (Figure 1)), the reanalysis mean from the most recent datasets, namely ERA-Interim, MERRA and JRA55.

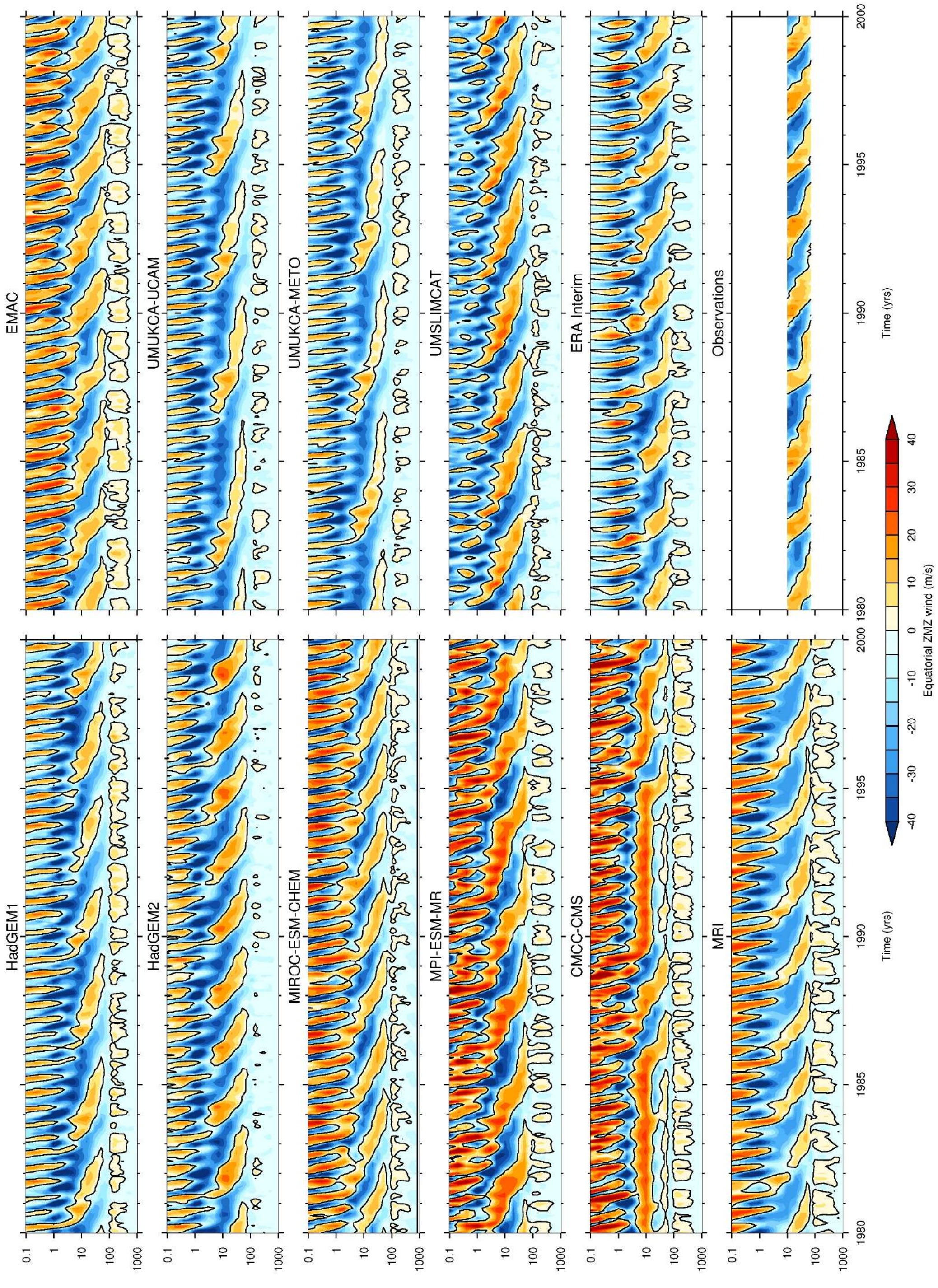

Figure 1. Equatorial zonal mean zonal wind time-height series from models and the ERA-Interim reanalysis, 1980-2000. Easterlies are blue, westerlies red. The zero wind line is shown in black. The observational dataset from Freie Universität Berlin (2015) is shown on the bottom right for levels 10-70hPa.

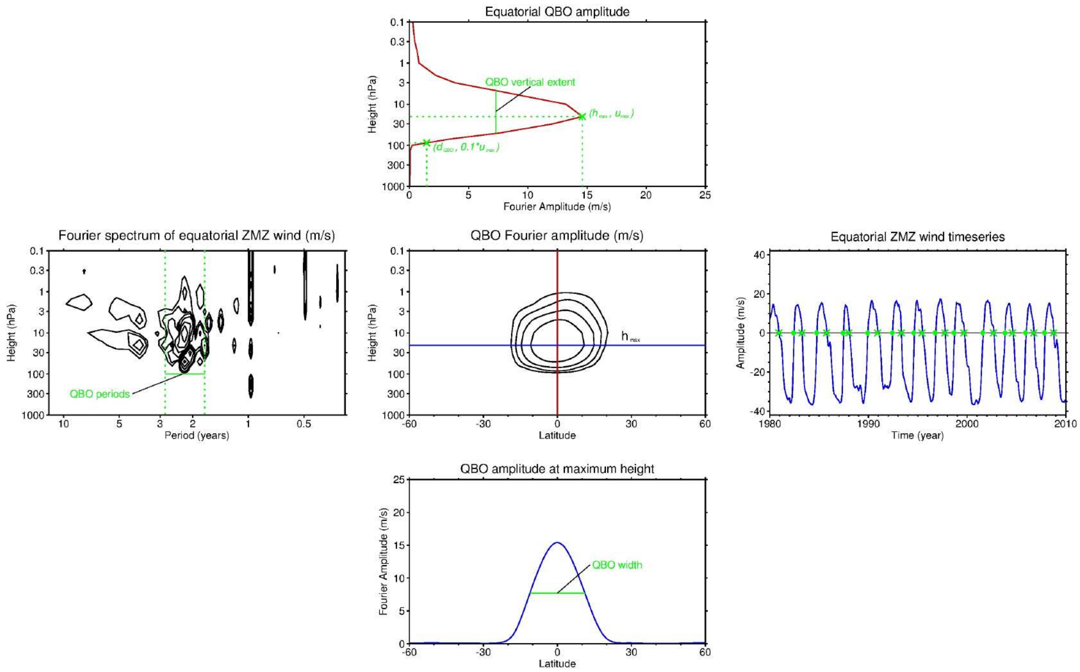

Figure 2. Derivation of QBO $\bar{u}$ characteristic metrics, exemplified with the reanalyses mean:

Middle row: Mean Fourier spectrum (left) of equatorial zonal mean zonal wind. Contours are drawn at 1, 2, 4, 8 and 16 m/s. The Fourier harmonics around 2 years are averaged to give the latitude-altitude QBO amplitude (middle, same contours). From the $\bar{u}$ timeseries at h_{max} (right), the period of each single QBO cycle is calculated and the easterly/westerly amplitudes are identified.

From the latitude-altitude QBO structure, a cross section at the equator (red) is taken to derive the QBO height profile (upper) and one at 20 hPa (blue) for the latitude profile (lower). From the height profile, the vertical extent, the depth d_{QBO} as well as the maximum Fourier amplitude (u_{max}) can be identified. The latitude cross section at h_{max}, where the equatorial QBO Fourier amplitude peaks, serves to define the latitudinal extent of the QBO.

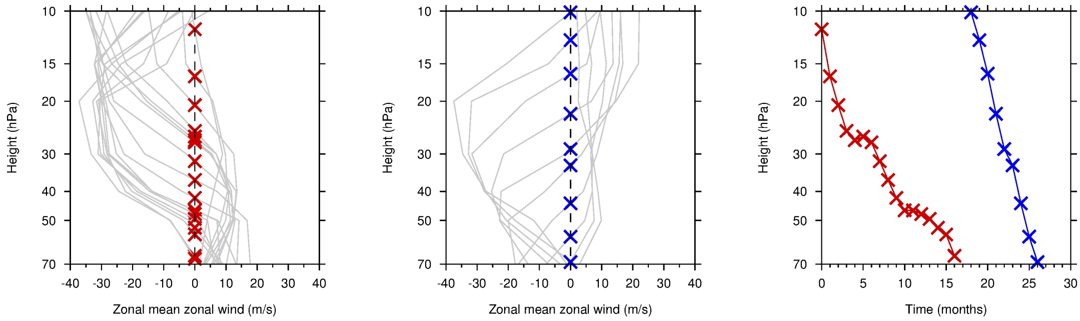

Figure 3. Equatorial $\bar{u}$ profiles in consecutive months for a descending easterly (left) and westerly (middle) shear zone from the FU Berlin observations (1964-1966 cycle). The heights of phase change in each month are shown in red/blue and are displayed in the right panel.

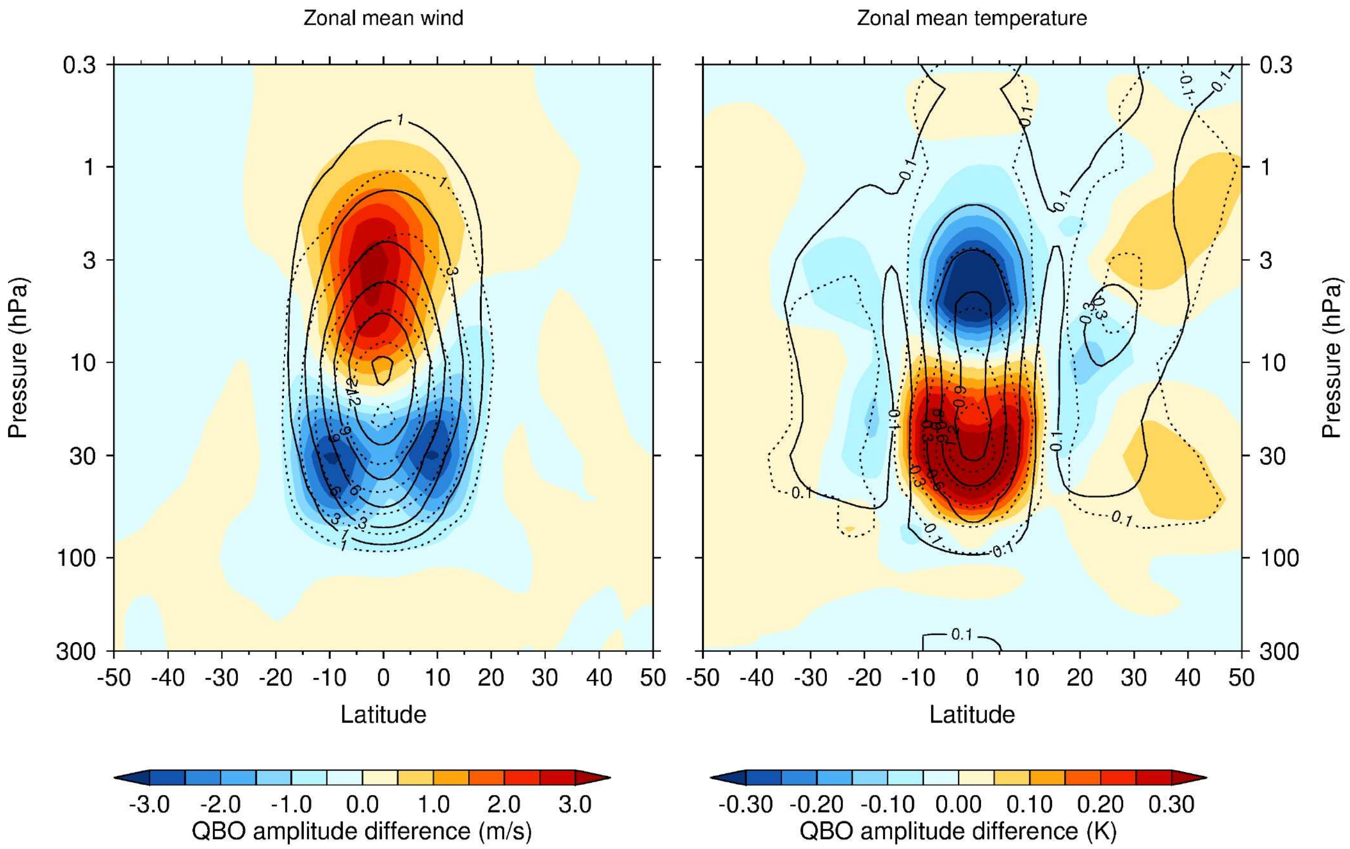

Figure 4. Comparison of QBO amplitudes in u (left) and T (right) from models (solid contours) and reanalyses (dotted contours). The colours show the difference models-reanalyses with blue depicting an underestimation by models and red an overestimation.

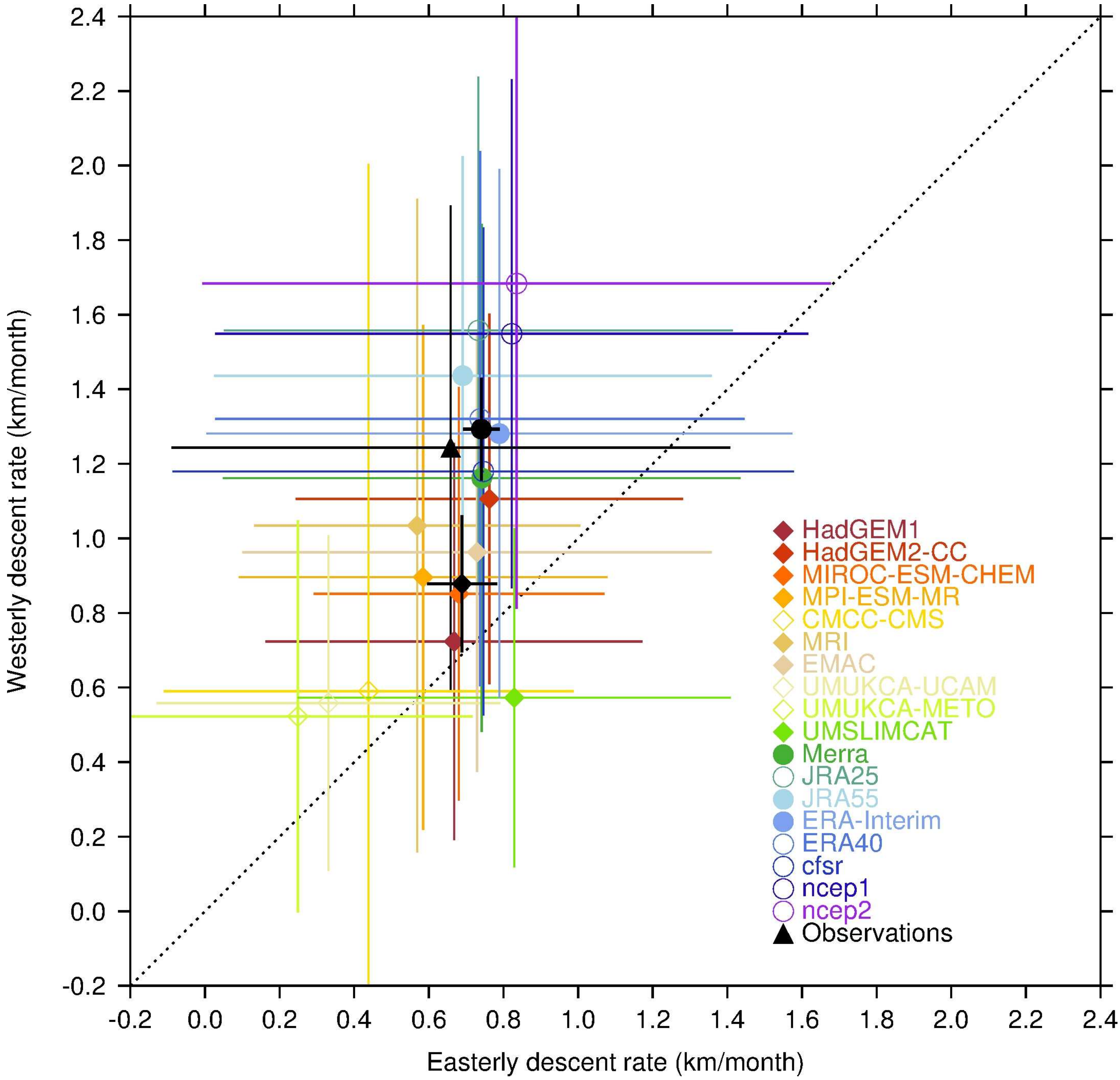

Figure 5. QBO easterly and westerly descent rates in models and reanalyses. The symbols (diamonds for models, circles for reanalyses and triangle for observations) show the mean and standard deviation within each dataset. The filled symbols contribute to the model/reanalysis mean as shown with the black diamond/circle. The dotted line represents equal descent rates for both shear zones as orientation.

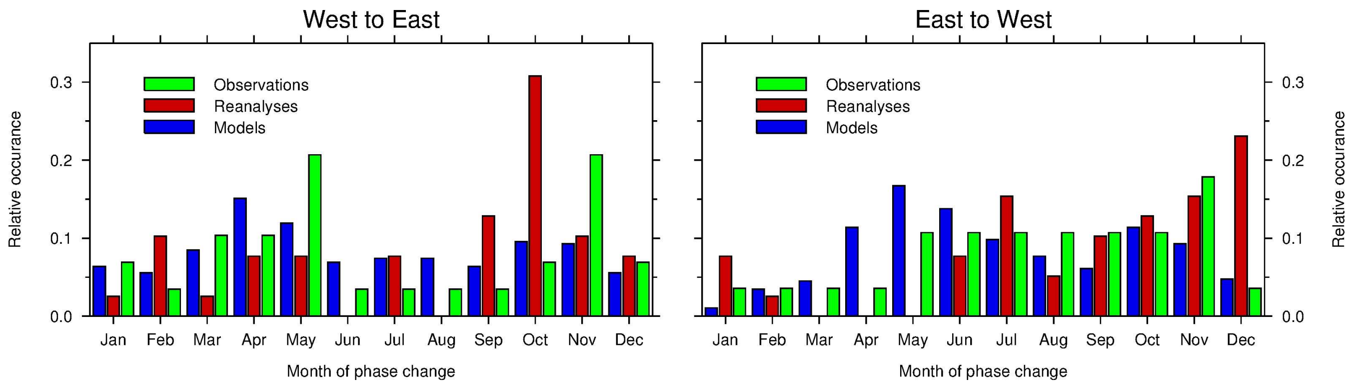

Figure 6. Timing of phase change in models (blue, excluding CMCC-CMS and both UMUKCA models), FUB observations (green) and reanalyses (red). There are 407/29/39 west-east changes (distribuation of relative occurence in left panel) and 411/28/39 east-west changes taken into account for models/observations/reanalyses.