Understanding the Local and Global Impacts of Model Physics Changes

Mark Rodwell and Thomas Jung

ECMWF
Shinfield Park, Reading RG2 9AX, United Kingdom

ABSTRACT

The real world and general circulation models are complex systems involving a large number of physical and dynamical processes. One useful approach to understanding these processes and for developing parametrizations is to use a hierarchy of simpler models. However, one must keep in mind that the fully complex system (which is used to produce our forecasts) may behave differently; for example due to interactions between processes and with the resolved flow. Hence there is a need to develop diagnostics that help us understand the physics, dynamics and interactions within a full general circulation model. This talk aims to introduce a few 'tools' that do precisely this. They are applied to cases where the physics of the model is changed in some way. The 'Initial Tendency' approach is used to help understand the local response to a given physics change, before interactions with the resolved flow have had time to occur. Diagnostics of equatorial waves (Kelvin, Rossby, Gravity, etc) and extratropical Rossby waves are used to help understand the global implications of a physics change.

1   Introduction

Localised tropical diabatic forcing anomalies can have an influence on the global circulation. The tropical response to localised heating has been discussed in terms of dynamical equatorial waves (Matsuno, 1966; Gill, 1980; Heckley and Gill, 1984). Midlatitude responses are often discussed in terms of 'teleconnection patterns' (Horel and Wallace, 1981; Hoskins and Karoly, 1981) . A knowledge of global teleconnections is essential for understanding the global climate and is useful for identifying remote 'causes' of observed seasonal-mean anomalies (Hoskins and Sardeshmukh, 1987).

Teleconnection patterns can be well simulated in models by imposing a prescribed (tropical) convective heating anomaly (Webster, 1972; Hoskins and Karoly, 1981; Hoskins and Rodwell, 1995; Greatbatch and Jung, 2007) or, more directly still, by imposing the upper-tropospheric divergence anomaly associated with convective outflow (Sardeshmukh and Hoskins, 1988). These studies have been essential to separate the dynamics of teleconnections from the physical mechanisms involved in their initiation. Of course, if extended-range predictions are to benefit from the existence of such teleconnectivity then these physical mechanisms also need to be well represented. But how do we efficiently improve the representation of physical processes when, in reality and in general circulation models, they are interacting so strongly with the resolved flow?

Here three approaches are brought together to gain a more complete understanding of the local and global impacts of a physics change within a full general circulation model. Firstly the 'Initial Tendency' methodology (Rodwell and Palmer, 2007), a development of that used by Klinker and Sardeshmukh (1992), is introduced and used. This methodology is able to separate the fast response of the physical processes from the subsequent interactions with the resolved flow. It thus offers a means of understanding the 'local' impact of a model physics change. A second approach uses equatorial wave theory to motivate the development of diagnostics that can aid in the understanding of the tropic-wide response to a given physics change. Finally, diagnostics are developed from the 'Rossby-wave source' approach of Sardeshmukh and Hoskins (1988). If a model change predominantly affects tropical physics, these diagnostics can help us identify and understand the subsequent extratropical stationary wave response.

The structure of this talk is as follows. In section 2, the datasets, the model and the experiments used in this talk are briefly outlined. In addition, a recent change in the model aerosol climatology is discussed. This aerosol change forms a useful case study for demonstrating the utility of the diagnostic tools that have been developed. In section 3, the mean global circulation for the June–August season is briefly discussed together with systematic climate errors in a recent version of the ECMWF model. The differences seen when the aerosol climatology is modified are also documented. Sections 4 and 5 introduce the Initial Tendency technique, show how it is closely related to the data assimilation process and how it can help in the assessment of model physics. Section 6 shows how Initial Tendencies can be used to assess climate prediction models. Section 7 shows how Initial Tendencies can be used to understand the local response to the aerosol change. To understand the tropic-wide response to a physics change, equatorial wave theory is introduced in section 8. In section 9, diagnostics motivated by equatorial wave theory are used to compare waves in general circulation models with those in reality. In section 10, the tropic-wide response to the aerosol change is understood in terms of the forcing of equatorial waves and diabatic feedbacks. In section 11, it is demonstrated that a tropical circulation change can lead rather directly to an extra-tropical divergent wind response. Vorticity and Rossby wave diagnostics are introduced and used to gain a better understanding of how the full extra-tropical circulation evolves to 'balance' this divergent wind change. The December – February season is be briefly discussed in section 12 and conclusions are given in section 13.

Because this talk is aimed at an audience of parametrization specialists, a pedagogical approach is taken to explaining the dynamical responses. Much of the content of this talk, with the exception of the equatorial wave sections, can be found in Rodwell and Palmer (2007), Rodwell and Jung (2008b) and Rodwell and Jung (2008a).

2 The Model, Data and Integrations

2.1 Observational data

Upper-air fields for the period 1962–2001 come from the ECMWF 40-year Re-Analysis dataset (ERA-40, Uppala et al., 2005). This dataset is derived using the 3-dimensional variational data assimilation system. The data assimilation process ingests data from almost all available sources. These include top-of-the-atmosphere radiative fluxes at many different wavelengths obtained from satellites as well as radiosonde ascents, drop-sondes and 'SYNOP' station reports.

Precipitation observations for the period 1980–1999 come from Xie and Arkin (1997).

Out-going long-wave radiation (OLR) measurements for the period 1990–2006 come from the National Oceanic and Atmospheric Administration (NOAA) satellite data (Liebmann and Smith, 1996).

2.2 Model description

A detailed description of the ECMWF model can be found at *http://www.ecmwf.int/research/ifsdocs/*. A brief overview of two aspects salient to the present talk: radiation and convection is given here. Note that more recent versions of the model, not used here, include updates to both the radiation and convection schemes.

The radiative heating rate is computed as the vertical divergence of the net radiation flux. Long-wave radiation is computed for 16 spectral intervals using the 'Rapid Radiation Transfer Model' (RRTM: Mlawer et al., 1997). The short-wave radiation part, which is computed for 6 spectral intervals, is a modified version of the scheme developed by Fouquart and Bonnel (1980). Since the computation of the radiative transfer equation is very expensive, the radiation scheme is ordinarily called at 3-hourly intervals and on a lower-resolution grid. Temporal and spatial interpolation are used to get these calculations onto the model grid. In some of the experiments discussed here (see below) it has been important to call the radiation scheme at every timestep.

Cumulus convection is parametrized by a bulk mass flux scheme which was originally described by Tiedtke (1989). The scheme considers deep, shallow and mid-level convection. Clouds are represented by a single pair of entraining/detraining plumes which describe updraught and downdraught processes.

The ECMWF model uses spherical harmonics to represent the prognostic fields. These harmonics are (triangularly) truncated at some total wavenumber, M. With the introduction of a two time-level semi-Lagrangian advection scheme in 1998, a linear, rather than quadratic, grid has been used for the calculation of physical tendencies. The triangular resolution is therefore defined as $T_L M$, and this equates approximately to a resolution in degrees of $180^o/M$ (the half wavelength of the shortest resolved zonal wave at the equator).

2.3 Aerosol changes

The primary case study of this talk is based around a recent change in model aerosol climatology. Here the 'old' and 'new' aerosols are discussed. Further details can be found in Rodwell and Jung (2008b) and references therein.

Until October 2003, the aerosol climatology used in the ECMWF operational forecasting model was based on that of Tanre et al. (1984). This climatology is specified as annual mean geographical distributions of various aerosol types: 'maritime', 'continental', 'urban', 'desert'. This aerosol climatology will be referred to here as the 'old aerosol'. Figure 1(a) shows the geographical distribution of the total optical depth for the old aerosol at 550 nm (an optical depth of d for a particular wavelength attenuates radiation at that wavelength by a factor e^{-d} as it passes through the atmosphere. This attenuation can be by scattering and absorption). The maximum optical depth (0.74) is seen to occur over the Sahara and this is dominated by desert (*i.e.* soil dust) aerosol.

In October 2003, a new aerosol climatology was implemented in the ECMWF forecast system (at cycle 26R3). This climatology is based on global maps of optical depths for a range of aerosol types compiled by Tegen et al. (1997). The aerosol types included are sea-salt, soil-dust, sulphate, organic carbon and black carbon. (Background stratospheric aerosol was left unchanged). Atmospheric loading for a given aerosol type is deduced from emission/transport modelling studies. This climatology will be referred to here as the 'new aerosol'.

For the new aerosol climatology in July (Figure 1(c)) the region of maximum total optical depth (maximising at a value of 1.05) is now located over the Somali Peninsula and out into the Arabian Sea associated with the transport of dust by the monsoonal Somali Jet. The aerosol optical depth over the Sahara is more than halved. The January aerosol of the new climatology (Figure 1(b)) also shows major differences with the old annual-mean climatology (Figure 1(a)). The magnitude of these changes is comparable with the uncertainties in mineral dust loadings summarised by Zender et al. (2004).

For the short-wave, in addition to being able to scatter radiation, some aerosol types such as soil-dust and black carbon can also absorb. By absorbing short-wave radiation, these aerosols can have a very direct impact on atmospheric temperatures. A measure of the relative strength of absorption is given by the 'single scattering albedo'. This is the ratio of scattering efficiency to total light extinction (scattering plus absorption). The single scattering albedo and other aerosol optical properties used within the ECMWF forecasting system are calculated following Hess et al. (1998). For desert aerosol, the single scattering albedo is around 0.888. For clean maritime air it is around 0.997. Since the differences between the new and old aerosol are particularly associated with desert aerosol, it is possible that absorption as well as scattering will be an important mechanism in the response.

In the ECMWF model, the aerosol concentration does not impact the cloud microphysics. Hence indirect aerosol effects such as how larger numbers of cloud condensation nuclei can lead to more, smaller and longer-lived cloud droplets and thus changes in the radiation budget are not represented. Instead, the local and global impacts of the change in aerosol climatology discussed here must arise from the direct effects of aerosol.

The last decade has seen further advances in aerosol estimation and, in this respect, the 'new' aerosol climatology cannot be considered as state-of-the-art. However, Tegen et al. (1997) show that comparisons with ground-based sun photometer measurements are reasonable and the 'new' aerosol climatology remains in ECMWF's
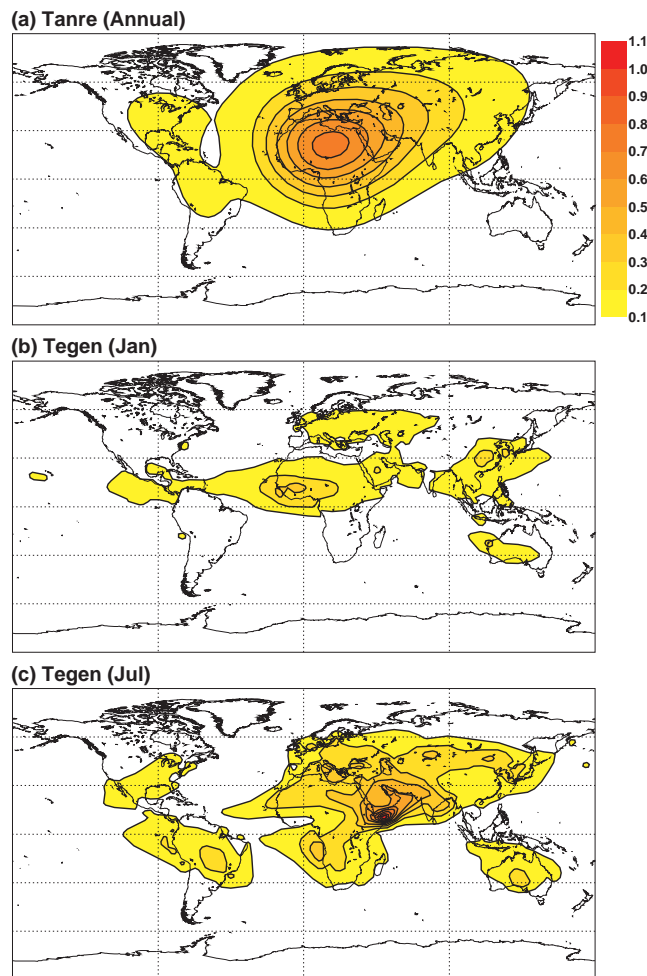
Figure 1: Optical depths at 550 nm associated with the model aerosol climatology. (a) The 'old' annually-fixed climatology of Tanre et al. (1984). (b) The 'new' January climatology of Tegen et al. (1997). (c) The 'new' July climatology of Tegen et al. (1997). The smallest contour is 0.1 and the contour interval is 0.1.

operational forecast model.

2.4 Seasonal integrations

To assess the climate of the atmospheric model and response of this climate to changes in model physics, sets of seasonal integrations have been made for 40 December–February and June–August seasons for the period 1962 to 2001. The initial conditions for these integrations are based on 1 April and 1 October analyses from the ERA-40 dataset, respectively (the first two months of each forecast were discarded). Sea-surface temperatures and sea-ice cover are also taken from ERA-40. These are based on monthly-mean values from the HadISST dataset (Rayner et al., 2003) up to November 1981 and weekly-mean values from the NOAA/NCEP 2D-Var dataset (Reynolds et al., 2002) thereafter. See figure captions for model cycle and resolution information.

It should be emphasised that the seasonal-mean climates and climate anomalies that will be shown are those from an atmospheric model in the presence of prescribed, realistic sea-surface temperatures. A good representation of the atmospheric processes included in this model is a pre-requisite for good atmosphere-ocean coupled simulations.

2.5 Weather forecasts

As will become apparent, it is difficult to isolate the direct effect of a particular model change in seasonal or climate simulations because this direct effect will be obscured by interactions and feedbacks with the resolved flow. The use of weather forecasts can greatly help in this regard because these are initiated from atmospheric states where the circulation is much closer to a real state of the atmosphere. In this study, sets of medium-range weather forecasts are started every 6 hours for the months of January and July 2004.

The weather forecasts are initialised-with, and verified-against, analyses produced by ECMWF's 4-dimensional variational data assimilation system (4DVAR, Rabier et al., 2000). 4DVAR starts with a 'first guess' from a previous model forecast and essentially involves iteratively nudging the non-linear and tangent-linear versions of the model to the new observations. Hence the analysis can be quite strongly dependent on the model used within the data assimilation. Since forecast errors and tendencies will be diagnosed at very short lead-times, a fair comparison of models (for example the models with new and old aerosol) requires that sets of analyses are produced: one set for each model.

These weather forecast integrations use model cycle 29R1 and are run at T_L159 ($\approx 1.1^o$) horizontal resolution with 60 levels in the vertical and a timestep of $\frac{1}{2}$ hour. In these forecasts, the radiation scheme is called every timestep with computations carried-out on a T_L63 linear grid.

2.6 Statistical testing

Where a statistical test, this refers to a two-sided Student's t-test of the difference of means. In every case shown here both distributions are based on the same set of dates and times and so a more powerful paired t-test is performed. Since autocorrelation could reduce the effective number of degrees of freedom in a timeseries, this is taken into account by using an auto-regressive model of order one (von Storch and Zwiers, 2001). A 'dual colour palette' has been developed to aid the display of statistical significance. This approach is used in some of the figures; with bolder colours indicating significant anomalies and the more pale colours indicating non-significance. Where confidence intervals are shown in plots, they are also based on the Student's t-distribution function and autocorrelation is taken into account in the same way. In general, an x% significance level can be thought of as a $(100-x)$% confidence level.

It is clearly important to always establish the statistical significance of any difference. One might be content to discover that an apparent improvement to the climate of the model is statistically significant but this talk is about going further. Here, the aim is to understand the reasons behind a climate improvement (or degradation).

3 June–August model climate response to the change in aerosol

Figure 2(a) shows mean June–August precipitation (shaded), low-level (925 hPa) wind vectors and 500 hPa geopotential height contours from the observational data. The summer monsoons of southern Asia, North Africa and Mexico, together with their associated low-level inflows are clearly evident. In the winter (southern) extratropics, a strong westerly jet is evident from the tightness of the geopotential height contours. In the summer (northern) hemisphere, the jet is weaker.

Figure 2(b) indicates the statistically significant mean errors for the model with the old aerosol. These errors include too much precipitation on the northern flank of the north African monsoon, strong wind biases over the sub-tropical north Atlantic and extratropical circulation biases to the south and southwest of South Africa. The effect of the change in aerosol (Figure 2(c); note the change in shading interval for precipitation) is a reduction in these particular mean errors so that they are no longer apparent in the mean errors with the new aerosol (Figure 2(d)). Elsewhere, mean errors are largely unchanged. The main degradation is perhaps the increased mean error in precipitation off the north-east coast of South America.
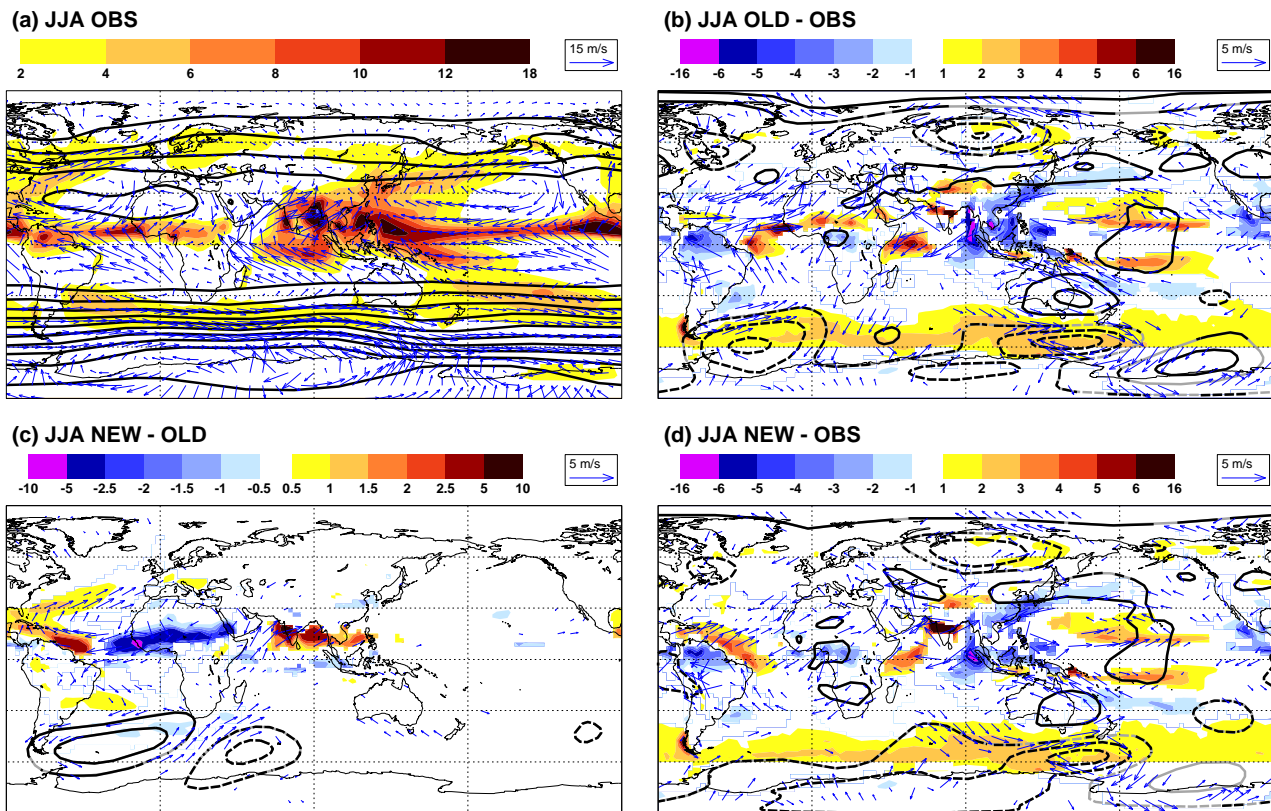
Figure 2: Diagnostics of June–August seasonal-mean total precipitation (shaded in mm day^{-1}), 925 hPa horizontal wind vectors (see scaling vector) and 500 hPa geopotential heights (see below for contour interval). Observational data come from Xie and Arkin (1997) for precipitation over the period 1980–1999 and from ERA-40 for the other fields over the period 1962–2001. Forecast data come from the 'seasonal integrations' covering the same period as for the observations. (a) Mean observed. (b) Mean model error with the 'old' aerosol. (c) Mean difference: model with 'new' aerosol minus model with 'old' aerosol. (d) Mean model error with the 'new' aerosol. Precipitation and wind differences are only plotted where seasonal-mean differences are statistically significant at the 10% level. Height differences are contoured solid for positive, dashed for negative, grey where not significant and with contour interval of 10 dam in (a) and 2 dam in (b)-(d). The seasonal integrations are made using model cycle 26R3 run at $T_L 95$ ($\approx 1.9^o$) horizontal resolution with 60 levels in the vertical and a timestep of 1 hour. The radiation scheme is called every three hours with computations made on a $T_L 95$ grid.

While seasonal-mean diagnostics can indicate changes and hopefully improvements in model climate, it is difficult to obtain a good understanding of how these changes come about. This is particularly the case when feedbacks are involved. To emphasise this point, we now digress from the aerosol example and look at mean forecast errors for a range of forecast lead-times.

4 Mean forecast error as a function of lead-time

Figure 3 shows 500 hPa temperature errors averaged over all operational 0 UTC forecasts made at ECMWF for the season December–February 2007/8. The four plots (a–d) show these mean errors for the forecast lead-times of 1, 2, 5, and 10 days, respectively.

At Day 1 (Fig. 3a), there is a uniform and statistically significant warm error over much of the tropics. (5% significance is indicated by the use of the bold colours, insignificance by the use of the pale colours). Generally there is also a cool error over the northern mid-latitudes. By Day 2 (Fig. 3b), the mean errors have got stronger

(note the change in shading interval) although there is no visible increase in the area that is statistically signifi-cant. Through Days 5 and 10 (Fig. 3c,d), the maximum values of mean errors continue to grow but the uniform pattern of tropical error seen at day 1 is replaced by a more complex pattern with a decreasing area over-which the mean error is statistically significant.

An interpretation of these results is that by days 5 and 10, interactions, teleconnections and loss of predictability have confused a simple investigation of the root causes for the mean forecast error. Statistical significance actually increases as the lead-time *decreases*. Taken to the ultimate extreme, one might expect that the best lead-time to use when searching for physical parametrization deficiencies would be at timestep 1 of the forecast! (see e.g., Klinker and Sardeshmukh, 1992). In fact timestep 1 introduces other problems associated with sampling the diurnal cycle so here the focus will be on the first few timesteps. These are, in fact, the timesteps within the data assimilation window and it is therefore appropriate to discuss data assimilation before moving to the 'Initial Tendency' methodology.

5   The data assimilation / forecast cycle

In the data assimilation process, the aim is to produce an 'analysis' that is as close to the observations as possible but also being (approximately) a valid model state. This analysis is then used as the initial conditions for a weather forecast. The data assimilation starts with a 'first guess' forecast initiated from a previous analysis. If one first assumes that the model used to make this first guess forecast is 'perfect', then the data assimilation / forecast cycle can be represented by the schematic diagram in Fig. 4a. The blue curve represents a timeseries of
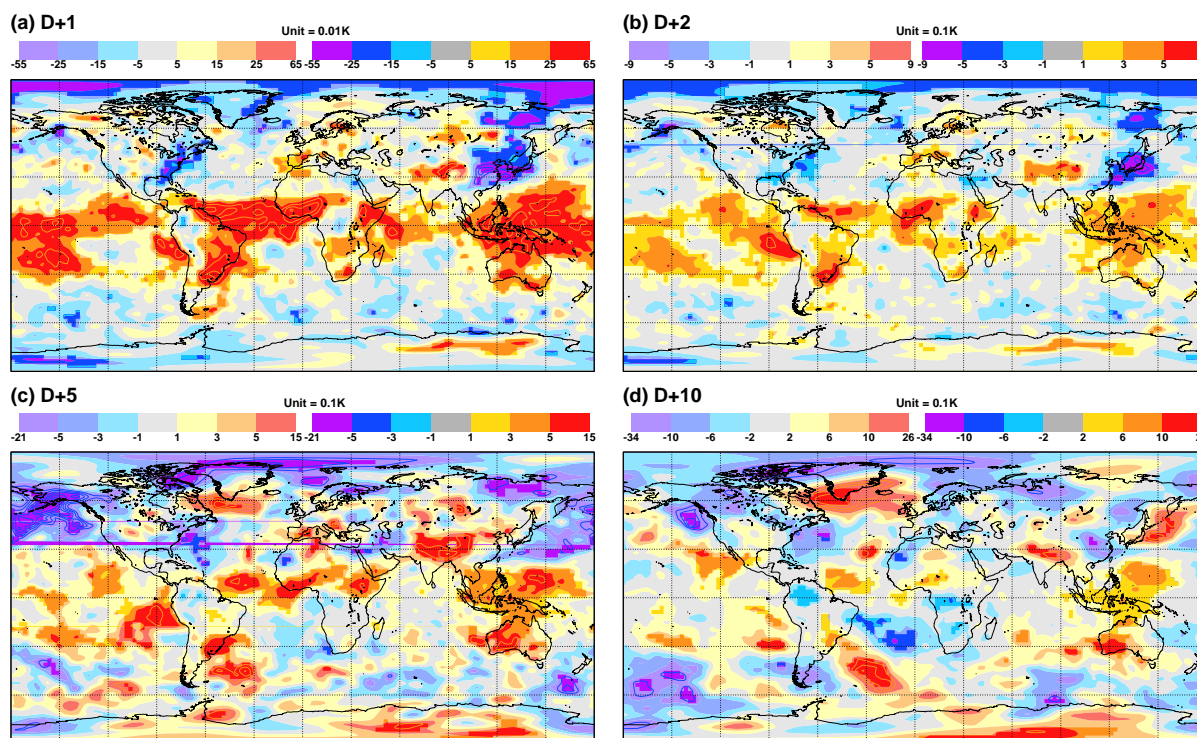

Figure 3: Mean forecast error for temperature at 500 hPa averaged over all forecasts initiated at 0 UTC and verifying within the season December–February 2007/8. The panels show the mean forecast error for a selection of forecast lead-times. (a) At a lead-time of one day (D+1). (b) D+2. (c) D+5. (d) D+10. Bold colours indicate that the mean forecast error is statistically significantly different from zero at a significance level of 5used to extend the colour shading scheme where necessary. The contour interval is the same as the shading interval.

observations at a given location. The red curves show successive first guess forecasts for four data assimilation / forecast cycles. The red dots represent successive analyses (based on prior observations) used to initialise the first guess forecasts. Chaos ensures that a first guess forecast will diverge from the subsequent observations even if the model is perfect. In the data assimilation process a tangent-linear version of the model is used iteratively to find a new model state that is closer to the new observations. The 'analysis increment' (shown by the black dotted lines) is the difference between the new analysis and the first guess forecast. As indicated in the schematic, to first-order a perfect model will produce as many erroneously cold first-guess forecasts as it will produce erroneously warm first-guess forecasts. Hence, the analysis increments for a perfect model will average to zero over sufficiently many data assimilation cycles. Note that this is true (to first order) even if the observations are not perfect as long as they are unbiased.

If a model is not perfect and has a systematic error (we will assume it has a cooling tendency as seen in the second schematic diagram, Fig. 4b) then, on average, the first guess will be colder than the observations. This will be reflected in a positive mean temperature analysis increment (depicted by the fact that all the dotted arrows point upwards). Indeed, it is clear that the mean analysis increment is equivalent to (minus) the mean initial tendency (in units of, e.g., K/cycle). How might such a systematic error arise? If the observations are sufficiently unbiased (see below) then one needs to look for problems within the model. The concept of 'radiative-convective equilibrium' embodies the idea that radiative processes act to destabilise the atmosphere (heat the surface and cool the mid-to-upper troposphere) and the convection induced by this destabilisation acts to restore balance by cooling the surface and heating the mid-to-upper troposphere. With this idealised concept in mind, either a convection scheme that is too weak (given the observed temperature and humidity profiles) or a radiation scheme that is too strong (given the observed conditions; as embodied by the analysis) would lead to a systematic initial net cooling of the mid-troposphere. Hence the mean initial tendency (or analysis increment) is a diagnostic that can quantify model physics error. Since this diagnostic is based on the first few timesteps of the forecast, it focuses on local sources of error and not on errors that develop as a result of subsequent interactions with the resolved dynamics. The advantage of Initial Tendencies over analysis increments is that Initial Tendencies can be broken-down into the component tendencies from each physical and dynamical process within the model. We can, for example, diagnose convective and radiative tendencies separately. Before making such a break-down of the initial tendencies, a specific example from the operational data assimilation system will be discussed. In particular, the important role played by the observations will be considered and the assumption that these observations are sufficiently unbiased will be tested.

Fig. 5a shows the analysis increments for 500 hPa temperature for the same December–February 2007/8 season as used for the forecast error results (Fig. 3). In the tropics, where the Day 1 forecast error indicated an erroneous warming by the model (Fig. 3a), the analysis increment shows a compensating cooling increment. Similar correspondence is apparent in the extratropical regions too.

Such temperature increments will only occur if there are supporting observations. These observations do not need to be direct observations of temperature since any observable quantity that can also be derived from the model state has the potential to influence the analysis. For example, one could consider as such a quantity the brightness temperature as observed by the "AIRS" infrared satellite channel 215. This brightness temperature represents a weighted mean of temperatures between about 700 hPa and 300 hPa; with the weight maximising at around 500 hPa. Using these weights, it is possible to derive the brightness temperature from the model state and thus make a comparison between the observed value and that predicted in the first guess forecast. In essence, the data assimilation processes iteratively modifies the model state in order to minimise the observation minus first guess difference for all such derived (and underived) quantities (subject to other constraints). Fig. 5b shows the mean observation minus first guess for this brightness temperature. The pattern agreement between the analysis minus first guess (Fig. 5a) and the observation minus first guess (Fig. 5b) indicates that AIRS channel 215 is one source of observations that 'support' the increments. Other obvious sources of supporting observations (not shown) include the "AMSUA" microwave channel 5 and the (somewhat sparse) tropical radiosonde network. Agreement between independent observation sources and consideration of the magnitude of likely observation biases leads to the conclusion that the increments are justified or, put another way, that the model really does have a tropical warming bias at 500 hPa. See Rodwell and Jung (2008a) for further discussion.

(b) Schematic diagram of data assimilation / forecast cycle (perfect model)

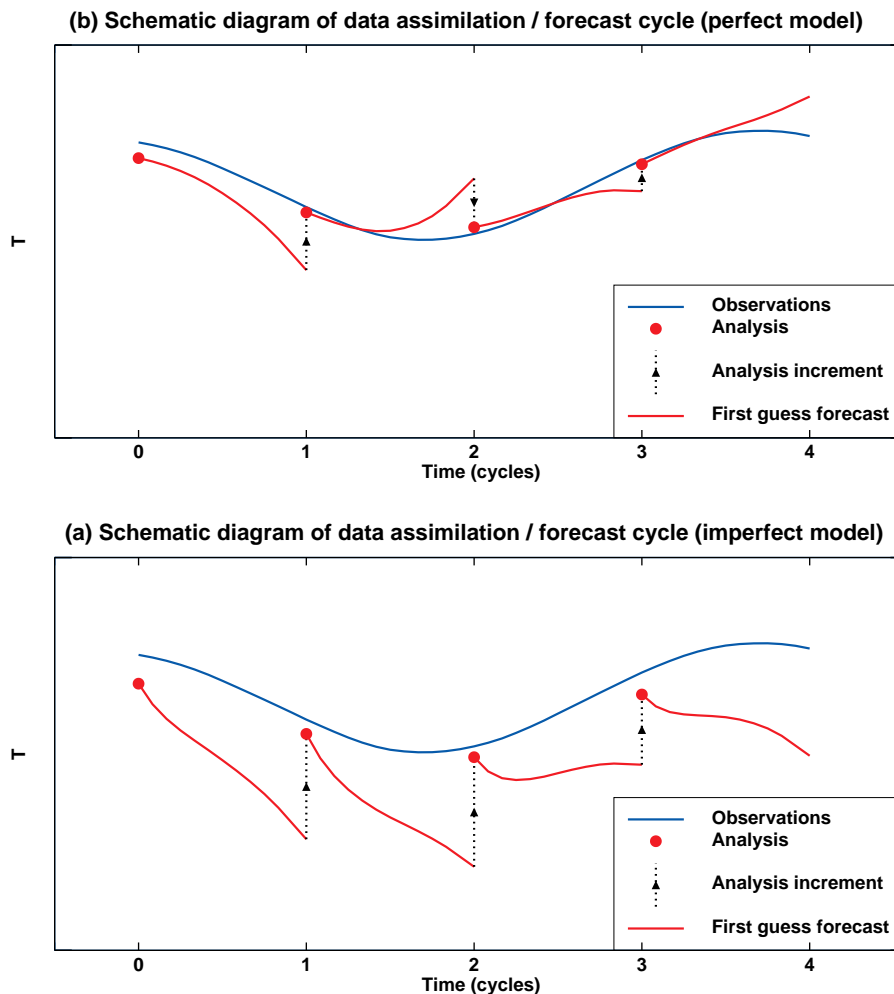**(a) Schematic diagram of data assimilation / forecast cycle (imperfect model)**

Figure 4: Schematic diagrams of the data assimilation / forecast cycling process for (a) a perfect model and (b) an imperfect model which suffers from systematic error. The blue curve represents an observed timeseries (e.g. of temperature at some specified location). These observations can contain random errors but these errors are assumed to be 'sufficiently' unbiased (when averaged over a large number of data assimilation cycles). The red dots represent successive analyses which used to initiate model forecasts. These forecasts, represented by the red curves, are used to supply "first guess" fields for the subsequent data assimilation cycle. An "analysis increment", represented by a black dotted line, shows how the incorporation of a fresh set of data draws the analysis closer to the observations. See the main text for further explanation.

6   Initial Tendencies: Assessment of climate prediction models

It has been shown that the mean analysis increment is the same as minus the mean initial tendency (if this tendency is averaged over the lead-time in the first guess forecast that corresponds to the validity time of the analysis). Under certain reasonable assumptions, a perfect model should lead to zero *mean* analysis increments and, therefore, zero mean initial tendencies. One can argue that *the smaller the mean initial tendency, the better the model*. This statement leads to a method of assessing and comparing models that has been used by Rodwell and Palmer (2007) to show one means of reducing uncertainty in climate change predictions. Below, we briefly discuss these results.

A major component of climate change uncertainty is associated with model uncertainty and an established method of sampling this model uncertainty is to produce 'perturbed model ensembles'; where sets of perturbations are made to tuneable parameters within a basis climate model (Murphy et al., 2004). With such a perturbed
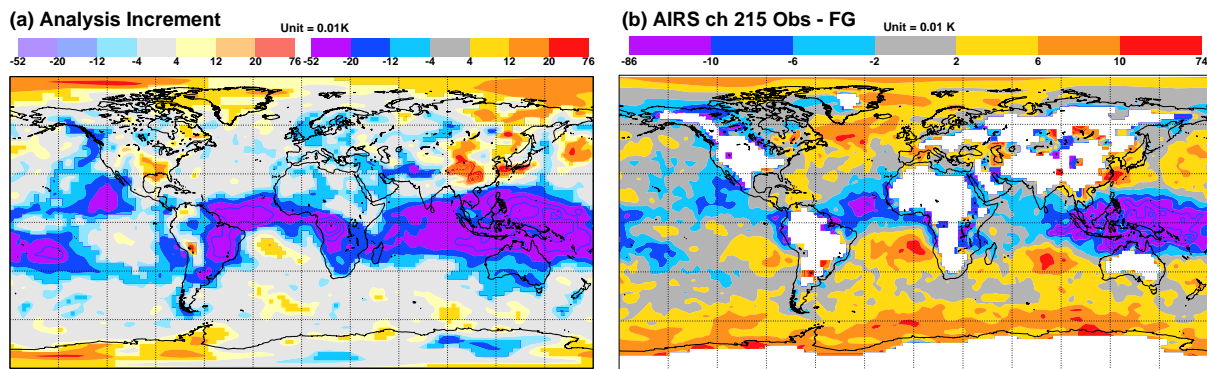
Figure 5: Mean results based on all 0 and 12 UTC data assimilation cycles within the season December–February 2007/8. (a) Mean analysis increment of temperature at 500 hPa. (b) Mean "first guess departure" (observation minus first guess forecast) for the "AIRS" satellite channel 215. The weighting function for this channel maximises at about 500 hPa.

model ensemble experiment, Stainforth et al. (2005) concluded that an 11K global-mean warming due to a doubling of atmospheric carbon dioxide was a real possibility. To make this assertion, thousands of perturbed climate models (based on the Hadley Centre's HADAM3 atmospheric model and incorporating a mixed-layer ocean) had to be calibrated and run for hundreds of years. Clearly, this is a computationally expensive exercise. Stainforth et al. (2005) found that it was a reduction in the 'convective entrainment parameter' that led to their most extreme global warming predictions.

By calculating initial tendencies for a single month of data assimilation cycles and very short forecasts, Rodwell and Palmer (2007) were able to demonstrate, with much less computational expense, that Stainforth et al.'s reduction in the convective entrainment parameter is unphysical (when applied to the ECMWF model at least). For example, the red curve in Fig. 6a shows the vertical profile of mean initial tendencies averaged over the Amazon/Brazil region. The data comes from the January 2004 data-assimilation / weather-forecast experiments using the control model detailed earlier. The error bars indicate 70% confidence intervals (see the figure caption for further details). The red curve in Fig. 6b shows the corresponding profile for the data-assimilation / weather-forecast experiment using the model with the reduced entrainment parameter. It is clear that the magnitudes of the mean initial tendencies are much greater in this region when the entrainment parameter is reduced. In effect, the first guess forecast for the reduced entrainment model diverges rapidly from the true state of the atmosphere and continually requires large analysis increments to put it back on course. Without these increments, the day 5 bias of the model (shown with the black curves) is seen to be up to 3K (at around 200 hPa).

Thinking in terms of a model timestep, the mean initial tendency is the sum of the tendencies produced by all the physical, dynamical and numerical processes within the model. Hence the mean initial tendency can be thought of as the 'Net mean initial tendency'; *i.e.* the sum of the mean initial tendencies from all the individual process. For simplicity, the word 'mean' will be dropped from now on. The tendencies associated with the more dominant processes are shown in Fig. 6. In the control model (Fig. 6a) it is evident that convective heating in this monsoonal region ('Con'; blue) is balanced by dynamical cooling due to ascent ('Dyn'; orange) and radiative cooling ('Rad'; green). In the concept of radiative-convective equilibrium, radiative cooling (*i.e.* destabilisation) of the atmosphere is thought to be balanced by convective warming (*i.e.* stabilisation) of the atmosphere but here it can be seen that dynamical processes are also highly important. Other important terms in the thermal budget are associated with vertical diffusion ('V.Dif'; brown, which includes sensible heating effects) and large-scale precipitation ('LSP'; pink).

What happens initially when the entrainment parameter is reduced is that heat and moisture are not detrained so quickly from a convective plume, the plume does not loose buoyancy so rapidly, and thus the convection strengthens and attains a higher altitude. Since the large-scale dynamical cooling is more strongly constrained by the observations, and responds more slowly than the convection, it cannot produce the initial cooling that

would be required to balance the increased convective heating. Consequently, the model is out-of-balance, its net initial tendencies are larger and the model is a much 'less perfect' model than the control! If the perturbed models of Murphy et al. (2004) and Stainforth et al. (2005) that incorporated this entrainment reduction were rejected or down-weighted then this would strongly reduce their uncertainty in climate change.

The Initial Tendency technique would appear to be a very powerful tool for assessing climate models. Note that it can only be used to assess processes (represented *or unrepresented* in the model) that act on short timescales. so other tests would still be required. Further benefits of the Initial Tendency technique, including linearity aspects, are discussed in Rodwell and Palmer (2007).

7   Initial Tendencies: Understanding the local impacts of the aerosol change

Fig. 7a shows the difference in initial tendencies (new minus old aerosol) for the North African region in July. This North African region is the area where the seasonal simulations displayed changes in mean monsoon precipitation (Fig. 2c). See the figure caption for more details. The question is, can these initial tendency differences be used to gain a better understanding of the local physics response to the aerosol change?
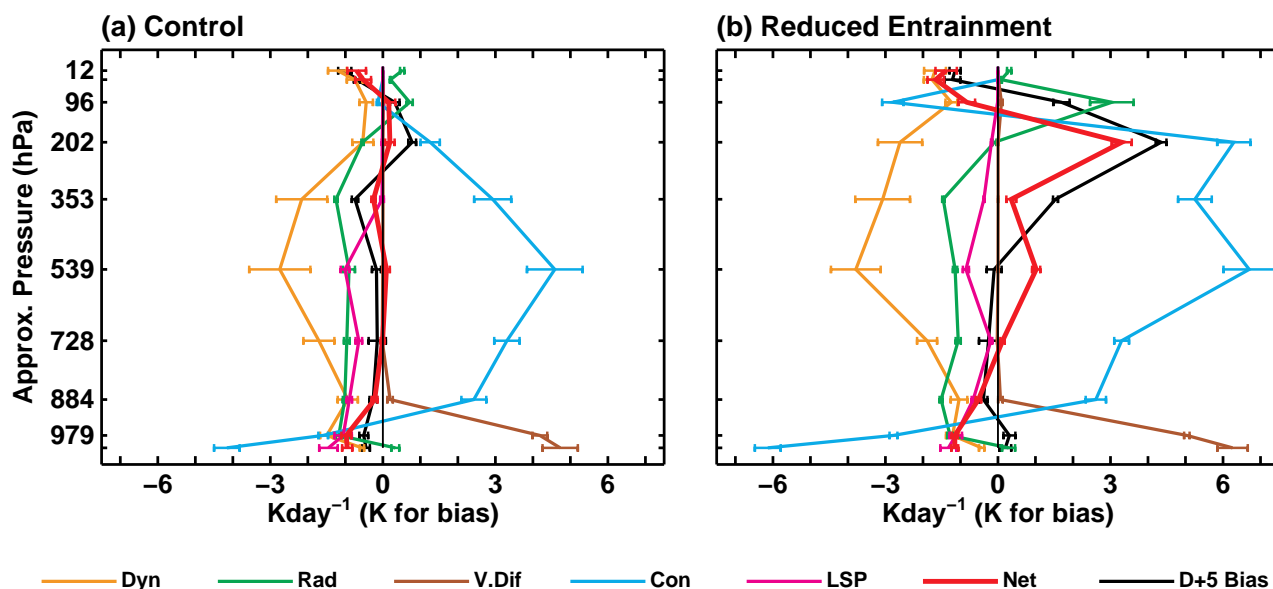

Figure 6: *Vertical profiles of initial temperature tendencies for the Amazon/Brazil region based on January 2005 'weather forecasts' for (a) the Control model (using model cycle 29R1) and (b) the same model but with reduced convective entrainment. The initial tendencies shown are indicated in the key and correspond to each model's dynamical tendencies, the tendencies from each of the dominant physical processes and the net tendency. Also shown are vertical profiles of the D+5 systematic error (bias). Mean tendencies are calculated on every 5th model level. The vertical coordinate is linear in pressure and represents the approximate pressure at these model levels. For each model version, a data assimilation / forecast cycle experiment was made with window length of six hours to generate analyses every 6 hours between 0 UTC on 27 December 2004 and 18 UTC on 26 January 2005 (and beyond). Five-day forecasts were started every six hours from these analyses to generate initial tendencies and D+5 biases. Initial tendencies for a given day are the tendencies accumulated over the forecast lead-times 0.5 to 6.5 hours and over the four forecasts made on that day. (Note that the verification times for a D+5 forecast correspond to the whole of January 2005 exactly). The bars shown indicate 70% confidence intervals based on the diurnally-integrated initial tendencies (summed over the four forecasts made each day). The Amazon/Brazil region (300^oE-320^oE, 20^oS-0^oN) was a region of systematic deficit in monsoon precipitation in seasonal integrations at model cycle 29R1. See the main text for more details.*

It was pointed out in section 1 that the change in aerosol was predominantly associated with a reduction in soil-dust. In addition, it was noted that soil-dust can absorb, as well as scatter, solar radiation. These effects are evident in the difference in the initial tendency due to radiation ('Rad'; green) which shows a cooling of the mid-to-lower troposphere. With less incoming short-wave radiation absorbed in the mid troposphere, more short-wave radiation reaches the surface. This is consistent with increased in near-surface heating by the radiation ('Rad'; green) and, in particular, by the vertical diffusion process ('V.Dif'; brown, which includes sensible heating effects). Hence the initial effect of the reduced soil-dust aerosol is an increase in radiative destabilisation of the atmospheric profile. The convection process ('Con'; blue) is seen to initially strengthen in response to this destabilisation. At first sight, this 'fast response' may seem perplexing since the aim is to explain a reduction in monsoon precipitation in the seasonal simulations.

To understand this apparent paradox better, the initial tendency differences can be compared with the differences in process tendencies later in the forecast. Fig. 7b shows the difference in tendencies at day 5 of the weather forecasts. The difference in radiative tendency ('Rad'; green) is broadly similar to the initial difference (Fig. 7a). However, the response to this radiative forcing change has evolved. It is possible that the semi-direct effect of aerosol (The term "semi-direct" is used to describe the mechanism whereby radiation absorption leads to warming and prevents condensation, Hansen et al., 1997) ensures that the initial convective heating anomaly (Fig. 7a 'Con'; blue) cannot balance the lower-tropospheric radiative cooling anomaly (Fig. 7a 'Rad'; green). Instead, a dynamical warming anomaly is required to complete the thermal balance. Since the dynamics responds more slowly, it is only later (for example at day 5) that this dynamical warming anomaly becomes established (Fig. 7b 'Dyn'; orange). Such dynamical warming is associated with low-level divergence and this has the effect of reducing the moisture convergence into the monsoon region and ultimately leads to a negative convective anomaly as seen in Fig. 7b ('Con'; blue).

Hence, in addition to the fast convective response to the aerosol change, there appears to be a slower mechanism involving interactions with the resolved flow. In effect, the reduction in soil-dust with the new aerosol prohibits an erroneous feedback with the resolved flow; whereby erroneous radiative heating within the aerosol layer leads to erroneous ascent, erroneous moisture convergence and erroneously strong convection.

If one were to examine the state of the model at any time beyond the initial tendencies, this slow feedback would dominate the signal and greatly hinder ones ability to understand the direct impact of the aerosol change. Importantly (but not shown here), the magnitudes of the net initial tendencies of temperature, moisture and winds are all improved with the introduction of the new aerosol. This reduction in net tendencies can be viewed as objective confirmation that the change in aerosol leads to a 'more perfect' model.

In terms of the present study's discussion of methodologies for model assessment, it is clear that the Initial Tendency approach provides a very powerful way of assessing model errors and model changes. Unlike top-of-the-atmosphere fluxes for example, they enable a 3-dimensional examination of the processes involved. They ensure that the observations are used in a consistent manner for both forecast initiation and forecast error assessment. Importantly, they allow the assessment to be made at atmospheric states close to reality, before many interactions and feedbacks have had time to take place. Referring back to the title of this section, Initial Tendencies allow a very *local* assessment of model physics. For example, the aerosol changes are predominantly over the north African region but there are differences further-a-field. These differences further-a-field (compare Figure 1a and c) will have impacts on the climate of long model simulations but will have minimal impact on the initial tendencies over north Africa. Because the Initial Tendency approach involves the production of analyses that are consistent with the model being tested, the Initial Tendency approach offers something extra to that offered by single column experiments, where the boundary conditions are considered to be independent of model version. For all these reasons, the Initial Tendency approach can provide developers of physical parametrizations with a very powerful method of assessing proposed model changes.
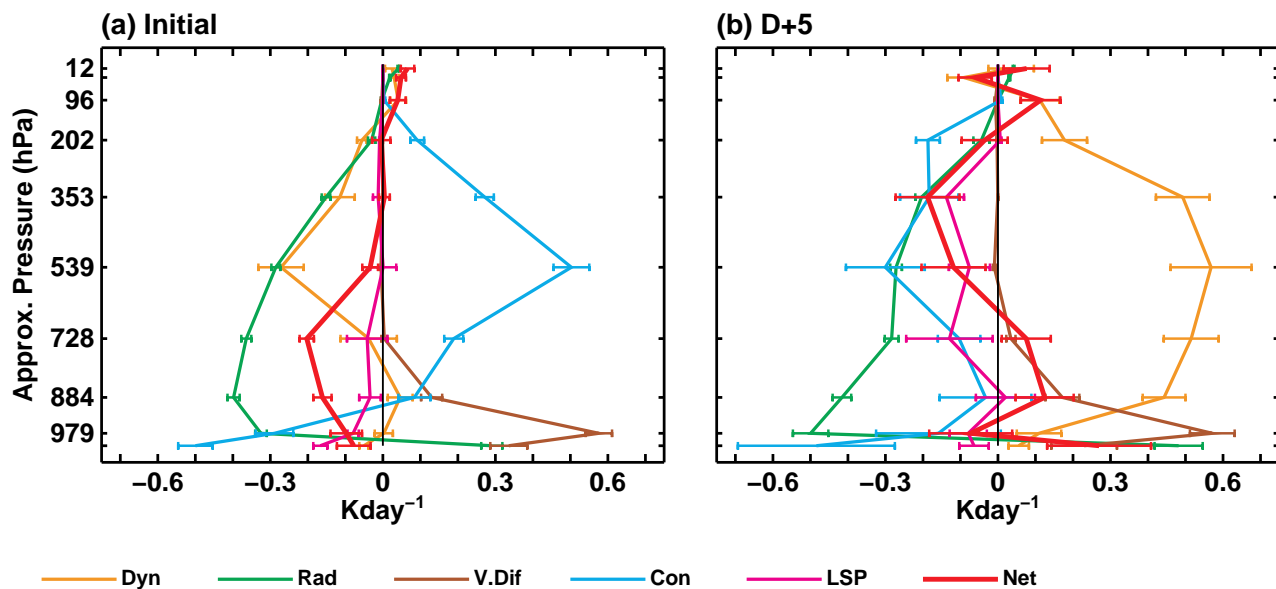
Figure 7: Vertical profiles of (a) initial tendency differences (new aerosol minus old aerosol) and (b) day 5 tendency differences of temperature averaged over the north African monsoon region using model cycle 29R1. The tendency differences shown are indicated in the key and correspond to each model's dynamical tendencies, the tendencies from each of the dominant physical processes and the net tendency. Results are based on 124 'weather forecasts' started every 6 hours from 26 June to 26 July 2004. Initial tendencies are accumulated over the forecast lead-times 0.5 to 6.5 hours and day 5 tendencies are accumulated over the forecast lead-times 120.5 to 126.5 hours. The north African region (20ºW to 40ºE, 5ºN–15ºN) corresponds to the region of decreased precipitation found in the 'seasonal integrations'. Other details are as in Fig. 6.

8   Equatorial waves: The Matsuno-Gill model

It has been shown that the direct radiative cooling effect of the aerosol change, and the strong erroneous precipitation feedback that it inhibits, lead to strongly reduced diabatic heating in the June–August season within the north African monsoon region. Two-layer shallow water equation studies on the linearised β-plane (Matsuno, 1966; Gill, 1980) (called here the "Matsuno-Gill model"; see below) demonstrate that such heating anomalies force equatorial waves that can communicate the response throughout the tropics. The aim in this section is to show how these equatorial waves arise, highlight their spatial and temporal characteristics, and demonstrate how a knowledge of these waves helps in the understanding of the tropic-wide response to a change in model physics.

To show some evidence for the existence of waves in the extra-tropics, Figure 8 shows Hovmöller diagrams of observed out-going long-wave radiation (OLR) for June–August 2006. The y-axis represents time and the x-axis represents longitude. Because OLR is strongly related to the cloud-top temperature, diagonally-oriented shading anomalies indicate longitudinally propagating waves of convection anomalies. In Fig. 8(a), OLR is averaged over the latitude band 5.0ºS– 5.0ºN. The wave highlighted by the black diagonal line could reflect the well-known, but poorly understood, Madden-Julian Oscillation (MJO) which tends to propagate eastwards around the equator. Observed MJO events generally propagate with a phase speed of about 10ms^{-1}, making a single rotation of the equator in around 30 to 60 days. Historically, models tend to produce too fast phase speeds for MJO-like features. This may be related to the faster phase speed of dry Kelvin waves (see later). In Fig. 8(b), OLR is averaged over the latitude band 7.5ºS– 17.5ºN. Numerous, mainly westward moving, waves are evident. The aim of this section is to develop a modelling framework that can explain some of the waves seen in the tropical regions. For this, the 'Matsuno-Gill' model is introduced.

The Matsuno-Gill model describes small perturbations from a state of rest in the equatorial region. Using
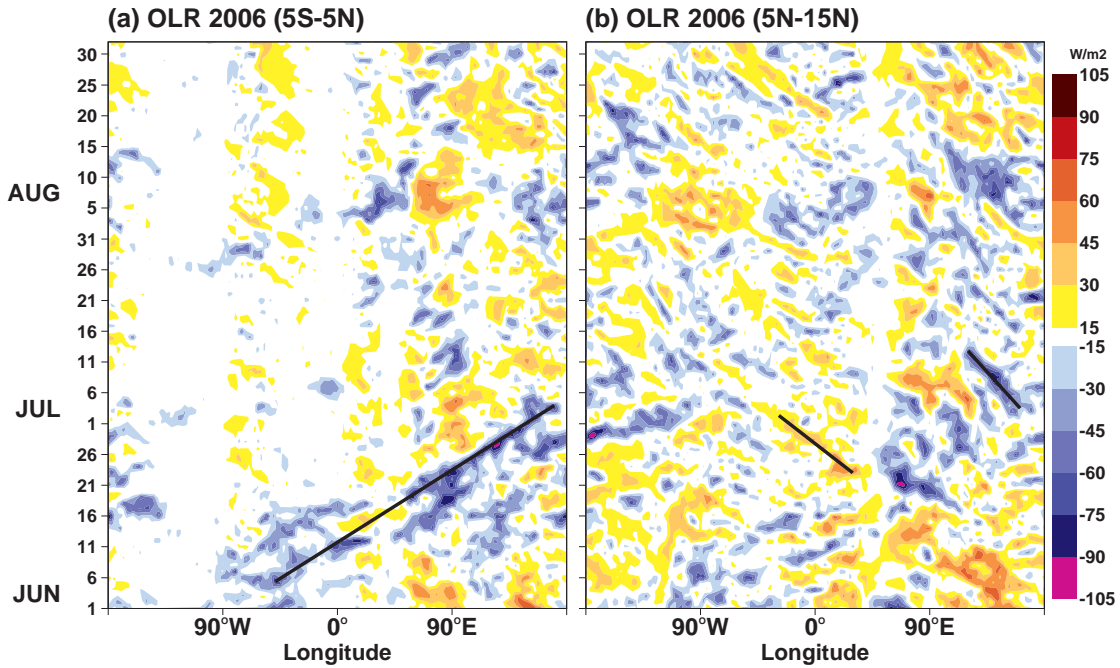
(a) OLR 2006 (5S-5N) (b) OLR 2006 (5N-15N)

Figure 8: Hovmöller diagrams of out-going long-wave radiation (OLR) using data from NOAA satellites. (a) Data averaged over the latitudinal band 5.0°S– 5.0°N. The black line indicates a possible Madden-Julian Oscillation (MJO) event with a region of anomalous convective activity (and thus low values of OLR) propagating eastwards.(b) Data averaged over the latitudinal band 7.5°S– 17.5°N. Black lines highlight a couple of westward propagating waves of convection anomalies, one positive, one negative.

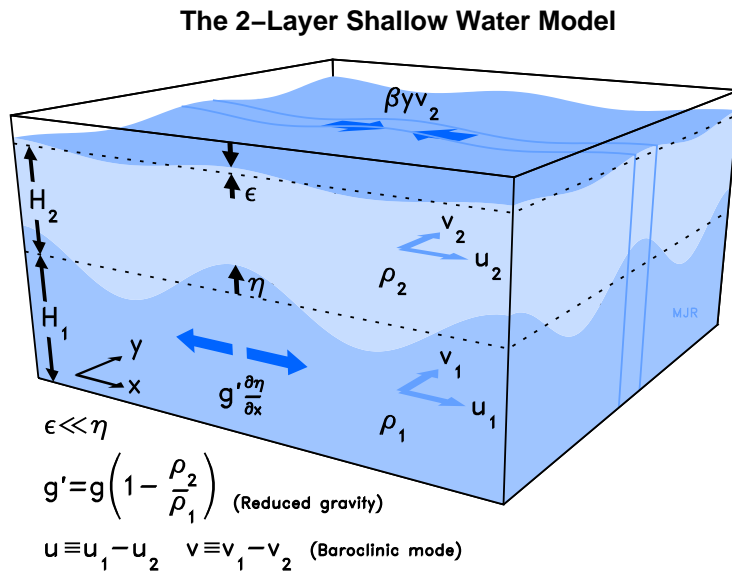## The 2–Layer Shallow Water Model

Figure 9: Schematic diagram showing the essential features of the 2-layer shallow water model as used by Matsuno (1966) and Gill (1980). See main text for more details.

this model, it is possible to gain insight into some of the observed waves in the tropics. The model is shown schematically in Figure 9. Two layers are essential to represent the tropical circulation where, for example, deep convective processes lead to opposite circulations in the lower tropospheric layer ($\mathbf{v}_1$ in layer 1) and upper tropospheric layer ($\mathbf{v}_2$ in layer 2). Indeed, the emphasis of the analysis here is on the waves in the baroclinic

component of the circulation, $\mathbf{v} \equiv \mathbf{v}_1 - \mathbf{v}_2$. The mean depths of layers 1 and 2 are H_1 and H_2 and their upper surface elevation anomalies are η $(\ll H_1)$ and ε $(\ll H_2)$, respectively. In general, $\varepsilon \ll \eta$.

The two layers have densities $\rho_1 > \rho_2$ and this difference in density allows horizontal pressure gradients to form and thus permits the representation of internal gravity waves. The baroclinic component of the horizontal pressure gradient can be written as $\rho_1 g' \nabla \eta$ where g' is the "reduced gravity". Gravity is "reduced" because the upper layer has a non-negligible density compared to that of the layer below (unlike the case of ocean surface waves, where the density of the air is negligible compared to that of water).

The model also includes planetary vorticity effects through the use of the "β-plane". β is the meridional gradient in planetary vorticity and its inclusion leads to Coriolis 'forces' of the form $-\beta y(\mathbf{k} \times \mathbf{v})$ where $\mathbf{k}$ is the unit vertical vector (see Fig. 9). It is this Coriolis effect that permits the model to represent internal Rossby waves. Rossby waves are discussed in more detail in the section 11.

The maths that leads to the free solutions of Matsuno-Gill model can be found partly in fluid dynamical text books (e.g. Gill, 1982) and partly in quantum physics text books (e.g. Landshoff and Metheral, 1979). The maths is presented below in (hopefully) a clear way for completeness.

The zonal and meridional momentum equations and the continuity equation for the baroclinic component of the flow can readily be written as

$$
\begin{aligned}
\frac{\partial u}{\partial t} - \beta y v + g' \frac{\partial \eta}{\partial x} &\approx 0 \\[2mm]
\frac{\partial v}{\partial t} + \beta y u + g' \frac{\partial \eta}{\partial y} &\approx 0 \\[2mm]
\frac{\partial \eta}{\partial t} + \frac{c_e^2}{g'} \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) &\approx 0 \quad,
\end{aligned}
\tag{1}
$$

where

$$
c_e^2 \equiv g' \frac{H_1 H_2}{H_1 + H_2} \equiv g H_e \qquad c_e = 20 \text{ to } 80 \text{ ms}^{-1} \quad.
\tag{2}
$$

Note that c_e is the propagation speed of a barotropic gravity wave in a single layer of depth H_e. Note also that there are no advection terms in equations (1) because the equations are linearised about a resting state. (However, the β term does imply the advection of planetary vorticity).

Solving for v, one obtains

$$
\frac{\partial}{\partial t} \left\{ \frac{\partial^2 v}{\partial t^2} + \beta^2 y^2 v - c_e^2 \left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} \right) \right\} - c_e^2 \beta \frac{\partial v}{\partial x} = 0 \quad.
\tag{3}
$$

One immediate solution to equation (3) is $v \equiv 0$. Inserting $v \equiv 0$ into equation (1), and looking for solutions which are separable in x and y and decay as $y \to \pm\infty$, one finds waves of the form

$$
u = u_0 e^{-\frac{1}{2}\beta y^2 / c_e} e^{ik(x - c_e t)} \quad.
\tag{4}
$$

Here, k is the zonal wavenumber of the wave and this can take any positive value. This class of waves is known as the equatorial Kelvin waves. Whatever the value of k, the wave propagates eastward with a speed c_e. Hence the Kelvin waves are non-dispersive with waves of different spatial scale all having the same eastward phase-speed. For reasonable values of H_1, H_2, ρ_1 and ρ_2, c_e may be between 20 and 80 ms^{-1}. One way to represent classes of waves graphically is to plot the angular frequency, ω, against zonal wavenumber, k. This is done
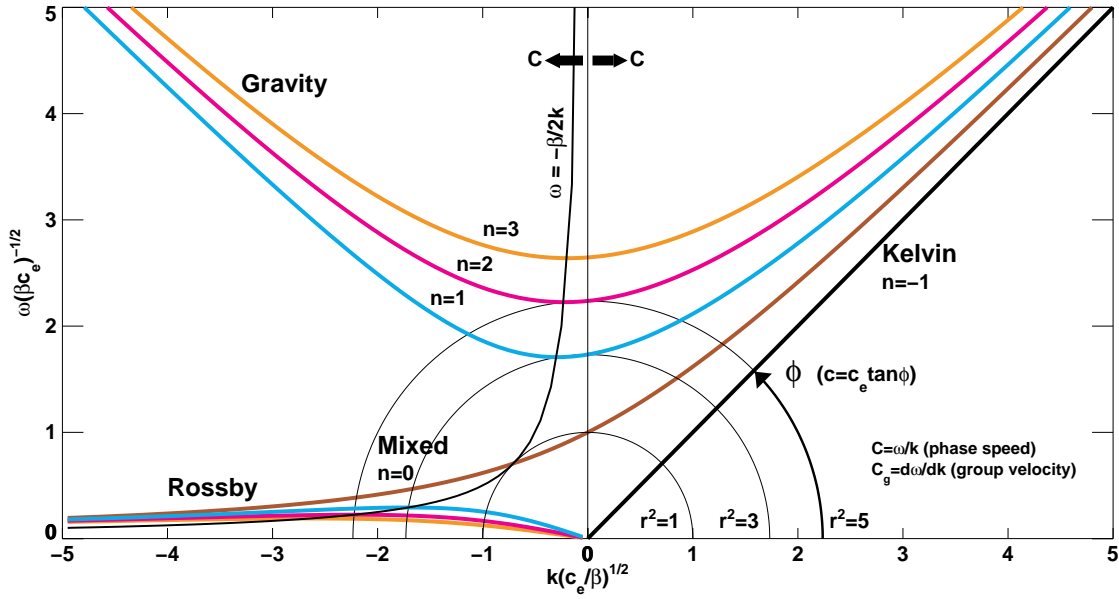
Equatorial Wave Dispersion Diagram

Figure 10: Dispersion diagram for equatorial waves. The vertical axis is angular frequency, non-dimensionalised by multiplying by the factor $(\beta c_e)^{-1/2}$. The horizontal axis is zonal wavenumber, non-dimensionalised by multiplying by the factor $(c_e/\beta)^{1/2}$. The black diagonal line represents the (eastward propagating) Kelvin waves, the brown curve, labelled $n = 0$, represents the mixed Rossby-gravity waves. The upper curves labelled $n = 1, 2, 3$ represent the first three gravity wave modes (eastward and westward propagating) and the similarly coloured lower curves represent the first three (westward propagating) Rossby waves for $n = 1, 2, 3$. Useful construction lines are also highlighted.

in Fig 10 (with both ω and k multiplied by non-dimensionalising factors). The Kelvin waves are depicted by the black diagonal line in Fig 10. The line is diagonal because the phase speed is $\omega/k = c_e$ for all k. (The other curves in Fig. 10 are discussed below). The structure of one such Kelvin wave is given in Fig. 11(b). Shading indicates η, the height perturbation of the layer interface, and the vectors indicate lower-layer winds. One feature of these Kelvin waves is that the zonal wind is in geostrophic balance with the meridional pressure gradient.

If $v \not\equiv 0$, then one can look for separable solutions to equation (3) which, as before, decay as $y \to \pm\infty$. Substituting $v = \hat{v}(y)e^{i(kx-\omega t)}$ into equation (3), one obtains the equation for the meridional structure, $\hat{v}(y)$, of the solutions:

$$\left(\frac{\beta^2}{c_e^2}y^2 - \frac{\partial^2}{\partial y^2}\right)\hat{v} = \left(\frac{\omega^2}{c_e^2} - k^2 - \frac{\beta k}{\omega}\right)\hat{v} \quad . \tag{5}$$

Equation (5) is actually Schrödinger's simple harmonic oscillator with the meridional structure, $\hat{v}$, being the Eigenvector and the multiplier of $\hat{v}$ on the right-hand-side being the corresponding Eigenvalue. It is easy to show that a solution (the simplest) is

$$\hat{v}_0 = e^{-\frac{\beta}{c_e}y^2/2} \quad , \quad \left(\frac{\omega^2}{c_e^2} - k^2 - \frac{\beta k}{\omega}\right) = \frac{\beta}{c_e} \quad (\equiv \lambda_0) \quad , \tag{6}$$

where $\hat{v}_0$ is an Eigenvector and λ_0 its corresponding Eigenvalue. In this solution, the meridional wind is maximum on the equator and decreases in strength as the latitude increases. As shown below, the other solutions to equation (5) can be obtained by induction from this first solution.
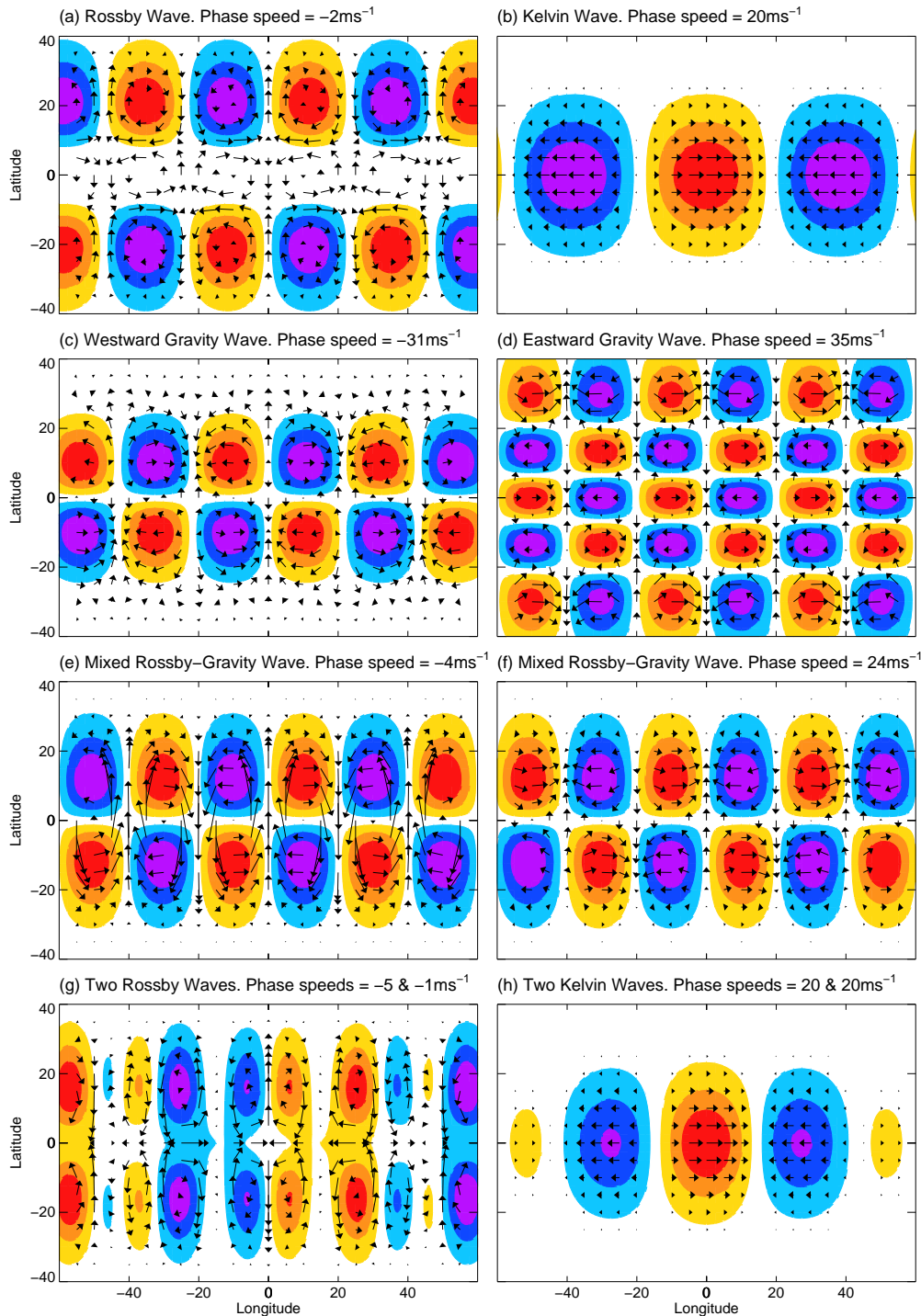
Figure 11: Example horizontal structures of a set of equatorial waves. Shading indicates the height pertur-bation of the interface between the two layers in the Matsuno-Gill model. Vectors show the horizontal winds in the lower layer. The wave-class (Kelvin, Rossby, Gravity, Mixed Rossby-Gravity) and the phase-speed of each wave are indicated above each panel. In (g), the superposition of two Rossby waves is shown. In (h), the superposition of two Kelvin waves is shown. The quoted phase speeds correspond to an assumed value of $c_e=20ms^{-1}$.

To demonstrate this induction, note that the differential operator on the left-hand-side of equation (5) can be

factored in two ways:

$$
\begin{aligned}
\left(\frac{\beta^2}{c_e^2} y^2 - \frac{\partial^2}{\partial y^2} \right) &= \left(\frac{\beta}{c_e} y - \frac{\partial}{\partial y} \right) \left(\frac{\beta}{c_e} y + \frac{\partial}{\partial y} \right) + \frac{\beta}{c_e} \\
&= \left(\frac{\beta}{c_e} y + \frac{\partial}{\partial y} \right) \left(\frac{\beta}{c_e} y - \frac{\partial}{\partial y} \right) - \frac{\beta}{c_e} \quad .
\end{aligned}
\tag{7}
$$

The extra terms at the ends of equation (7) come from the chain-rule for differentiation. Using equation (7), it is straightforward to show that, if $(\hat{v}_n, \lambda_n)$ is an Eigenvector / Eigenvalue solution then so is

$$
\hat{v}_{n+1} \equiv \left(\frac{\beta}{c_e} y - \frac{\partial}{\partial y} \right) \hat{v}_n \quad , \quad \lambda_{n+1} \equiv \lambda_n + \frac{2\beta}{c_e} \quad .
\tag{8}
$$

Hence we get an infinite series of solutions:

$$
\hat{v}_n = \left(\frac{\beta}{c_e} y - \frac{\partial}{\partial y} \right)^n e^{-\frac{\beta}{c_e} y^2 / 2} \quad , \quad \lambda_n = (2n+1) \frac{\beta}{c_e} \qquad \text{for } n = 0, 1, 2, \dots \quad .
\tag{9}
$$

The Eigenvectors, $\{\hat{v}_n\}$, are the Hermite polynomials (multiplied by $e^{-\frac{\beta}{c_e} y^2 / 2}$). The first few (for $n = 0, 1, 2, 3 \dots$) are given by

$$
\hat{v}(\tilde{y}) = \left\{ \begin{array}{l} 1 \\ 2\tilde{y} \\ 4\tilde{y}^2 - 2 \\ 8\tilde{y}^3 - 12 \\ \vdots \end{array} \right\} e^{-\tilde{y}^2 / 2} \quad , \quad \text{(where } \tilde{y} \equiv (\beta/c_e)^{1/2} y \text{)} \quad .
\tag{10}
$$

Each polynomial has one more node (latitudes where $\hat{v}_n$ is zero) than the previous polynomial and successive polynomials alternate between being asymmetric and symmetric about the equator (for symmetric modes $\hat{v}(-y) = -\hat{v}(y)$).

Using equations (5) and (9), the Eigenvalues $\{\lambda_n\}$ define the 'dispersion relation':

$$
\lambda_n \equiv \left(\frac{\omega^2}{c_e^2} - k^2 - \frac{\beta k}{\omega} \right) = (2n+1) \frac{\beta}{c_e} \quad .
\tag{11}
$$

Remember that k is the zonal wavenumber (so the wavelength is $2\pi/k$) and ω is the angular frequency (so the period of oscillation at any given point is $2\pi/\omega$). The dispersion relation shows how the wavelength and frequency of a wave are related to each other (and dependent on the value of n). The dispersion relation is cubic in ω and so, in general, there are three values of ω for any value of k and n. Only two roots are valid if $n = 0$ (see below). First, the roots for $n \geq 0$ are discussed.

For large ω, the $\beta k/\omega$ term in equation (11) is relatively unimportant and equation (11) becomes a quadratic with two roots. These two roots describe fast (eastward and westward propagating) equatorially-trapped gravity waves. The gravity waves (for $n = 1, 2, 3$) are represented in Fig 10 by the upper curves. Note that the westward solution is shown best with k negative and ω positive although it could also be shown with k positive and ω negative. Fig. 11(c) shows the structure of a westward propagating gravity wave. The phase-speed is -31ms^{-1}. In this case, there are two latitudes where $v \equiv 0$ and so this wave lies on the left-hand side of the $n = 2$ gravity wave curve in Fig 10. Fig. 11(d) shows the structure of an eastward propagating gravity wave with phase-speed 35ms^{-1} and $n = 3$.

For given k, the third root in equation (11) occurs with relatively small values of ω, when the $\beta k/\omega$ term is important. This root therefore corresponds to the slower (westward propagating) equatorially-trapped Rossby waves. The first three Rossby wave modes are indicated in Fig 10. Incidentally, when Fig 10 is plotted with the appropriate aspect ratio, the angle ϕ can be used to calculate the phase-speed, c, of any wave with the formula $c = c_e tan(\phi)$. For the Rossby waves, $c_e tan(\phi)$ is clearly negative and, compared to the gravity modes, small in magnitude. Fig. 11(a) shows the structure of a (westward propagating) Rossby wave with $n = 2$. The phase-speed for this wave is -2ms^{-1}. With a much slower phase-speed than the gravity modes, the winds associated with this Rossby mode are more in geostrophic balance with the height anomalies (the winds tend to rotate around the height anomalies).

When $n = 0$, the two valid roots for each value of k correspond to the class of 'mixed Rossby-gravity waves'. These mixed waves are indicated by the brown curve in Fig 10. Roots with westward phase-speeds behave a bit like Rossby waves. Fig. 11(e) shows such a westward propagating mixed Rossby-gravity wave. It has a phase-speed of about -4ms^{-1} and, since $n = 0$, there are no latitudes where $v \equiv 0$. These characteristics, together with the fact that $\beta y = 0$ at the equator, mean that these waves can be associated with strong cross-equatorial flow anomalies (as shown in Fig. 11e). Eastward roots tend to have higher phase-speeds and behave more like gravity waves. The structure of one such wave is shown in Fig. 11(f). The third root with $n = 0$ leads to an unbounded solution which is unacceptable.

Fig. 11(h) shows the super-position of two Kelvin waves. The non-dispersive nature of Kelvin waves means that the spatial pattern of the super-position does not change; it simply moves to the east. On the other hand, Fig. 11(g) shows the super-position of two Rossby waves. These waves have differing phase-speeds and so the spatial pattern of the super-position evolves with time. Note that the *group velocity* of such a wave packet can be eastward even though the waves are individually propagating westward.

9   Equatorial waves: Comparing models with observations

The dispersion diagram (as in Fig 10) can be used as the basis for assessing how well the simple Matsuno-Gill model 'predicts' waves in the fully complex atmosphere. It can also provide a means for comparing the waves simulated by complex models and those in the observations.

Figure 12, following Wheeler and Kiladis (1999), shows the activity (power) in out-going long-wave radiation (OLR) as a function of zonal wavenumber and frequency based on the December–February season for the years 1990–2005. The data has been first written as the sum of two components: one symmetric about the equator and the other asymmetric about the equator (for symmetric modes $OLR(-y) = OLR(y)$). 12(a) shows the power in the symmetric component of the observed OLR. Over-laid on this power spectrum are the theoretical dispersion lines for symmetric waves. There appears to be reasonable agreement between the theoretical curves and the regions of highest OLR wave activity. For example, a peak in wave power follows the diagonal line of the (eastward propagating) Kelvin waves. There is also enhanced power in the (westward propagating) Rossby wave regime. Note that the dispersion curves have been drawn using $c_e = 20 \text{ ms}^{-1}$. With this value of c_e, $\omega(\beta c_e)^{-1/2} = 1$ implies a period of 3.4 days (0.3 cycles per day) and $k(c_e/\beta)^{1/2} = 1$ relates to a wavelength of 6000km (7 cycles around the globe). Fig. 12 focuses on a more limited range of zonal wavenumbers and frequencies compared to Fig 10 because the observations are based on twice-daily sampling and waves with shorter wavelengths or shorter periods are not adequately resolvable.

In Fig. 12(a) there is also strong eastward wave power with very lower frequency (corresponding to a period of around 30 days). This power is associated with the Madden-Julian Oscillation (MJO). The fact that this power does not readily lie on one of the dispersion curves suggests that the MJO involves physics (such as convective coupling) which is not represented within the simple Matsuno-Gill model.

Fig. 12(b) shows the power in the asymmetric component of the observed OLR. Again, the theoretical dispersion lines for asymmetric waves are over-laid. Here, there is evidence of enhanced power associated with the mixed
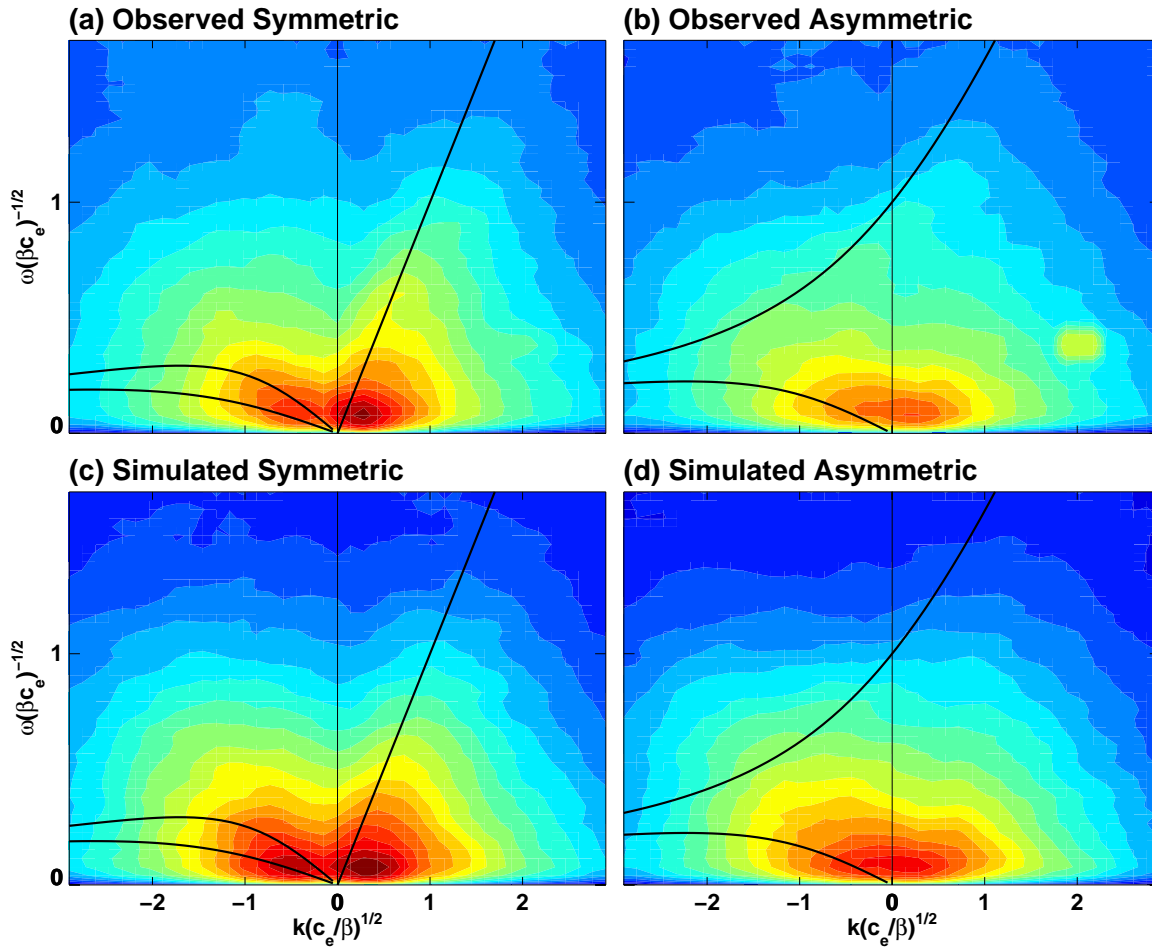
Figure 12: Power-spectra of the activity in out-going long-wave radiation as a function of zonal wavenumber and frequency based on the December–February season for the years 1990–2005. (a) For waves in the component of NOAA observed OLR variability that is symmetric about the equator (i.e. n odd). (b) As (a) but for the observed asymmetric component of variability (n even). (c,d) as (a,b) but based on simulated OLR variability in seasonal integrations of ECMWF atmospheric model cycle 32R3, run at resolution $T_L 159, L91$.

Rossby-gravity waves and the asymmetric Rossby modes. There also appears to be some asymmetric wave power at MJO temporal and spatial scales. The 'yellow blob' at $k(c_e/\beta)^{1/2} = 2$ is artificial and associated with aliasing of the data.

Fig. 12(c and d) show corresponding wave power from seasonal simulations of the ECMWF atmospheric model cycle 32R3. Convective parametrization changes associated with this model cycle led to changes in the wave power spectrum. It can be seen that there is reasonable agreement between the observed and modelled waves. However, there is too much low frequency wave power and the mixed Rossby-gravity waves are perhaps not so well distinguished.

10 Equatorial waves: Time-mean response to heating

Gill (1980) solved the equatorial wave equations as solutions to fixed heating in the presence of a linear damping term. The steady response can be understood in terms of the free wave solutions discussed above. Figure 13 shows Gill's 'monsoon' result. The red contours show the centre of off-equatorial monsoon heating. The flow to

the east, which has no meridional component, is clearly associated with a 'down-welling' Kelvin wave solution (compare with Fig. 11b). The wave is 'down-welling' in the sense that some of the ascent within the monsoon heating region (red contours) is balanced by descent within the Kelvin wave solution to the east (blue contours). In this idealised example, the flow to the west of the heating is the sum of two Rossby wave modes, one of which is the asymmetric mode highlighted in Fig. 11(a). Descent is also seen within the western part of the Rossby wave component of the solution. See Rodwell and Hoskins (1996) for further discussion on the reasons for this descent in the absence of the strong damping employed by Gill (1980).

If tropical convection is poorly simulated (for example, if there are systematic errors in the Asian monsoon) it is clear from these results that, quite quickly, the action of equatorial waves will lead to the development of errors in the large-scale flow. The vertical motions associated with these errors can trigger (or inhibit; Rodwell and Hoskins, 1996) remote convection errors. Faced with an erroneous seasonal-mean model climate it will, therefore, be very difficult to isolate the root cause of the error. A possible solution to this problem is to look at errors very early on in the forecast, before such interactions with the resolved flow have taken place. This was the main justification for using the Initial Tendency methodology discussed above.

From this equatorial wave theory, one would anticipate that the weakening of the north African monsoon due to the reduction in soil-dust aerosol would force non-dispersive, eastward-propagating Kelvin wave anomalies. In the seasonal-mean, a signature of these waves would be anomalous upwelling over the Indian Ocean. The substantially increased rainfall seen in Figure 2(c) over the northern Indian Ocean / Asian monsoon region (5mm day^{-1} over the west coast of India) is consistent with a triggering of convection by these Kelvin waves. If this is the case, this again highlights how feedbacks with the physics are able to enhance the dynamical forcing.

Equatorial wave theory also tells us that the cooling anomaly within the north African monsoon region and the heating anomaly over the northern Indian Ocean will force equatorial Rossby waves. The strengthened cross-equatorial and southwesterly low-level flow over the Arabian Sea in Figure 2(c), for example, is associated with the equatorial Rossby-wave response to the Asian monsoon heating anomaly (as in Rodwell and Hoskins, 1995). Similarly, the low-level wind anomaly over the sub-tropical Atlantic in Figure 2(c) is consistent with the Rossby wave response to reduced north African monsoon heating (Rodwell and Hoskins, 2001).

Hence it would appear that simple equatorial wave theory, together with the notion of coupling with physical processes such as convection, is able to explain most of the tropic-wide response to the aerosol change.
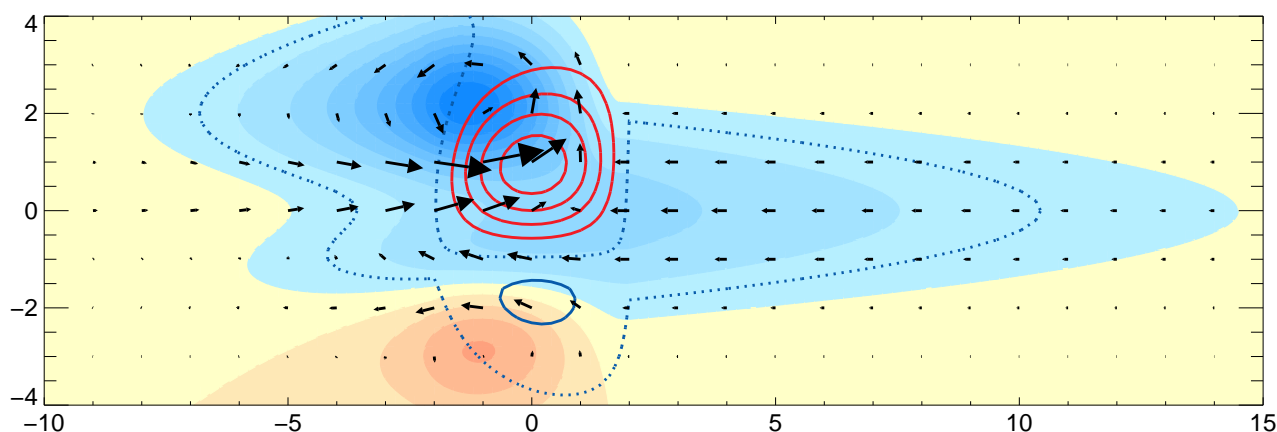

Figure 13: The steady solution, following Gill (1980), of the forced two-layer shallow water equations for the case of a 'monsoon-type' heating anomaly. Shading shows anomalous surface pressure (blue negative, orange positive), vectors show lower-tropospheric winds and contours show vertical motion (red ascent, blue descent, the zero contour is not plotted, the dotted blue contour is an extra vertical motion contour to emphasise the descent aspect of the solution).

11   Extratropical impacts: Rossby-wave forcing

Using the Initial Tendency analysis, it has been possible to explain the June–August local physical response to the change in aerosol. The tropic-wide response has been discussed in terms of equatorial wave theory and the likely enhancement by the diabatic physics. One feature in Figure 2(c) remains to be examined. This is the June–August Southern Hemisphere extratropical response which appears as an equivalent barotropic anticyclone–cyclone pair centred to the south of South Africa, with strong southwesterly winds in-between. At 500hPa, this extratropical feature appears disconnected from the tropical changes further north. Two-layer shallow water theory, which was used above to interpret tropical, internal, baroclinic waves, is not well suited to explaining this extratropical, external, equivalent-barotropic response. Instead, it is well known that Rossby-wave dynamics in the upper-troposphere provide the tropical-extratropical link for such a response. Rossby waves are associated with vorticity anomalies and so it is appropriate to start with a discussion of the vorticity equation. As with the equatorial wave theory, this discussion is presented in a pedagogical manner before it is related to some original results.

Vorticity is the curl of the wind. In 2D horizontal flow, it can be expressed as $\zeta_z = \partial v/\partial x - \partial u/\partial y \equiv \mathbf{k} \cdot \mathbf{\nabla}_z \times \mathbf{v}$, where $\mathbf{k}$ is the unit vertical vector and $\mathbf{\nabla}_z \times$ is the horizontal curl operator. For motions that rotate anticlockwise when viewed from above, $\partial v/\partial x$ is, in general, positive and $\partial u/\partial y$ is, in general, negative so the vorticity is positive. To understand the ways in which vorticity can change as the flow evolves and interacts with the physics, one can look at the curl of the momentum equations. The curl of the 3D momentum equations in an absolute frame of reference is given below along with some diagrams that indicate the meaning of each of the terms.

$$\frac{d\boldsymbol{\zeta}}{dt} \quad = \quad -\boldsymbol{\zeta}(\mathbf{\nabla} \cdot \mathbf{u}) \quad + \quad (\boldsymbol{\zeta} \cdot \mathbf{\nabla})\mathbf{u} \quad - \quad \frac{1}{\rho^2}\mathbf{\nabla}\rho \times \mathbf{\nabla}p \quad + \quad \mathbf{\nabla} \times \mathbf{F_u} \quad ,$$

Lagrangian   **Divergence**   **Tilting**   **Baroclinic**   **Friction**

(12)

Here $\mathbf{u}$ is the 3D wind, $\boldsymbol{\zeta} \equiv \mathbf{\nabla} \times \mathbf{u}$ is the absolute vorticity, ρ is density, p is pressure, and $\mathbf{F_u}$ is friction. In the diagrams accompanying equation (12), the vorticity vector is represented by an arrow. The blue arrow is the initial vorticity of a parcel of air (of unit mass) and the corresponding black arrow the same air parcel's vorticity some time later. The Lagrangian tendency (the term on the left-hand side) is the difference between these two vorticity vectors as one follows the trajectory of the air parcel. The divergence term describes the change in magnitude (not direction) of the vorticity. It expresses the concept of conservation of angular momentum. The example often given is the increase in rotation of an ice skater as they bring their arms in towards their body. Convergence decreases the moment of inertia of the skater and so the rotation increases in order to conserve angular momentum. The tilting term is fairly self explanatory and describes the change in direction (not magnitude) of the vorticity vector as the parcel is tilted. Note that there are also "stretching" components to both the divergence and tilting terms as written in equation (12) but these cancel. In this regard, the term "stretching", which is sometimes used to signify the spin-up of vorticity, can be a little misleading. However, unless a parcel's density changes markedly, stretching must be accompanied by convergence in the plane perpendicular to the axis of rotation. It is this convergence that results in the spin-up of vorticity, as indicated in the figure. The baroclinic term is the curl of the pressure gradient force. It leads to a tendency in vorticity if the pressure gradient and density gradient are not parallel. The thickness of the blue circle in the diagram for the baroclinic term is drawn to be proportional to the air density. The pressure gradient will accelerate the air with lower density more than the air with higher density and lead to a tendency in the vorticity. The frictional term is fairly self explanatory. If friction tends to retard the flow then this will lead to a reduction in the magnitude of the vorticity. There

is no gravitational component in equation (12) because the curl of the gravitational force is zero. Importantly, equation (12) also holds if **u** and the Lagrangian time-derivative are expressed relative to the rotating planet.

Making the "Shallow Atmosphere Approximations" (as done, for example, in the ECMWF model) and neglecting the vertical advection, tilting, baroclinic and frictional terms on scaling arguments for midlatitude synoptic systems, one arrives at the equation:

$$\frac{\partial \zeta}{\partial t} + \mathbf{v} \cdot \nabla \zeta = -\zeta (\nabla \cdot \mathbf{v}) \quad , \tag{13}$$

for the vertical component of absolute vorticity, $\zeta \equiv f + \mathbf{k} \cdot \nabla \times \mathbf{v}$, where **v** is the horizontal wind, **k** is the local unit vertical vector and f is the Coriolis parameter.

Traditionally (see, e.g. Sardeshmukh and Hoskins, 1988) the wind field is separated into divergent and rotational components, $\mathbf{v} = \mathbf{v}_\chi + \mathbf{v}_\psi$ (where $\mathbf{v}_\chi = \nabla \chi$, the wind component parallel to the gradient of the velocity potential, χ, and $\mathbf{v}_\psi = \mathbf{k} \times \nabla \psi$, the wind component parallel to the streamfunction, ψ). The components of equation (13) that are dependent on the divergent flow, $\mathbf{v}_\chi$, are moved to the right-hand side of the vorticity equation and regarded as a forcing from the tropics (associated for example with convective out-flow changes forced by sea-surface temperature anomalies or, as in the main example here, with aerosol changes). The remaining components, which are purely associated with the rotational flow, $\mathbf{v}_\psi$, are regarded as representing the extratropical barotropic response

$$\frac{\partial \zeta}{\partial t} + \mathbf{v}_\psi \cdot \nabla \zeta = -\nabla \cdot (\mathbf{v}_\chi \zeta) \quad . \tag{14}$$

To emphasise the traditional separation into 'tropical forcing' and 'extratropical response', the right-hand side of equation (14) is sometimes known as the 'Rossby-wave source'. It combines the divergence component and the component associated with advection by the divergent wind.

The vectors plotted in Fig. 14 show the time-mean upper tropospheric divergent flow response to the change in aerosol climatology, deduced from the seasonal integrations. These vectors highlight the anomalous upper-tropospheric convergence associated with the weaker north African monsoon and also the increased divergence over southern Asia and the northern Indian Ocean (which were associated earlier with coupling between the convection and the upwelling Kelvin-waves). Note that, for general circulation models, the height of convective outflow can be very sensitive to changes in model formulation and the height of convective outflow also varies from one tropically convective region to another. Hence, in order to obtain robust results from general circulation model output, it has been found to be useful to mass-average all upper-tropospheric diagnostics between 300 and 100 hPa. The divergent wind anomaly field associated with the tropical convection response to the aerosol change, Fig. 14, is seen to extend into the midlatitudes, where equation (14) suggests the potential for an influence on (or at least an interaction with) the extratropical vorticity budget. This influence would occur via changes in the Rossby wave source term which is also shown (shaded) in Fig. 14. The Rossby wave source is deduced using daily data from the seasonal integrations. The strongest Rossby wave source changes occur outside the tropics (over northern Africa and south of South Africa) where the divergent wind anomalies coincide with larger magnitudes in absolute vorticity. (The background absolute vorticity is shown with thin grey contours). The largest changes in the Rossby wave source are associated with changes in the divergent flow, not changes in the absolute vorticity.

The thick contours in Fig. 14 show the mean upper-tropospheric streamfunction change. There is a strong 'quadrapole' anomaly centred around Africa with low streamfunction anomalies over the subtropical north Atlantic and southern Indian Ocean and high anomalies over the subtropical south Atlantic and Arabian Peninsular (and Capsian Sea). This quadrapole is part of the equatorial wave solution to the aerosol forcing change. It strongly eliminates the mean errors in upper-tropospheric streamfunction in this region relative to the ERA-40 climatology (not shown). Notice also the wave-like features within the extratropical jet regions; in particular at around 50^oS south of South Africa and the wavy nature of the anomalous streamfunction contour at around
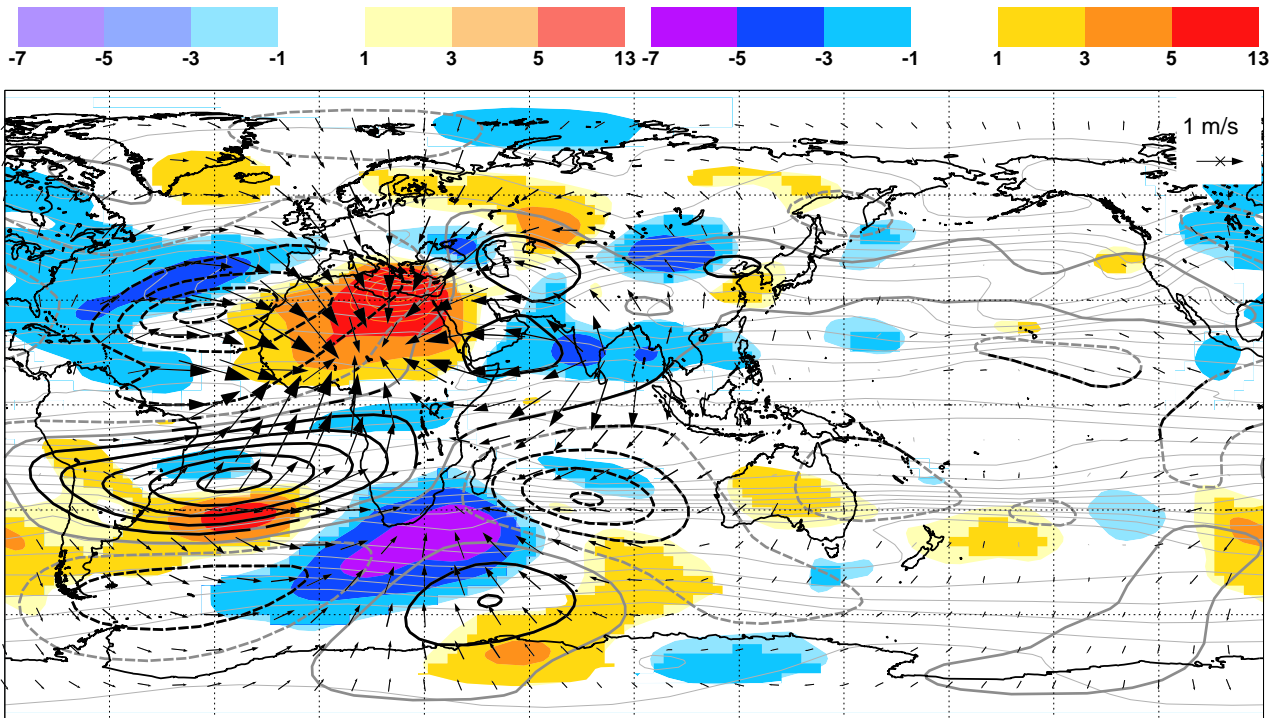
Figure 14: June–August mean change in upper tropospheric flow diagnostics from the 40 years of 'seasonal integrations'. Arrows show the change in divergent winds. Shading shows the change in Rossby-wave source derived from daily data. Thick contours show the change in streamfunction. The change is in the sense of 'new' minus 'old' aerosol. Also shown in thin grey contours is the mean absolute vorticity (the full field is shown; not the change). The diagnostics are derived from the two sets of 'seasonal integrations' (see main text for details). Daily data at 12 UTC is used throughout. The shading interval for the anomalous Rossby-wave source is generally $2 \times 10^{-11} s^{-2}$ but note that orange contours are used with the same interval to divide the most extreme (red) colour. The anomalous streamfunction contours are shown at ± 1, ± 3, ± 5,... $\times 10^6$ $m^2 s^{-2}$. Contours of the mean absolute vorticity are displayed at ± 5, ± 10, ± 15,... $\times 10^{-6} s^{-1}$. All quantities are vertically integrated between 300 and 100hPa. Black arrows, black contours and bold shading indicate 10% statistical significance for differences of seasonal-means.

45^oN over the North Pacific (this latter signal is not strictly significant at the 10% level and that is why the thick contour is grey). To examine how the extratropical time-mean rotational flow anomaly develops in response to the divergent flow changes, we look at the time-mean balances in equation (14). Since synoptic and intraseasonal variability does not appear to be important for the time-mean vorticity balance in the aerosol example (see below) and the time-mean of the time derivative can be neglected in these long simulations, the vorticity balance can be written as

$$ -\overline{\overline{\mathbf{v}}_{\psi}^{(N-O)} \cdot \boldsymbol{\nabla} \overline{\zeta}^{(O)}} - \overline{\overline{\mathbf{v}}_{\psi}^{(N)} \cdot \boldsymbol{\nabla} \overline{\zeta}^{(N-O)}} - \overline{\boldsymbol{\nabla} \cdot (\overline{\mathbf{v}}_{\chi} \overline{\zeta})}^{(N-O)} \approx 0 \quad , \tag{15} $$

where an overbar indicates a seasonal-mean over 92 days and a second overbar indicates a mean over the 40 years of simulations. The superscripts $^{(N)}$, $^{(O)}$ and $^{(N-O)}$ refer to the new aerosol, old aerosol, and new minus old aerosol, respectively. The advection of vorticity by the rotational flow has been decomposed into two parts; one associated with changes in the rotational flow and one associated with changes in the absolute vorticity.

Figure 15(a) shows the mean change in the Rossby wave source deduced from seasonal-mean anomalies (third term in equation (15). The similarity with the Rossby wave source deduced from the daily data (shaded in Fig. 14) emphasises that transient (intraseasonal) terms do not contribute greatly to the time-mean vorticity budget (see Rodwell and Jung, 2008b, for more discussion). The question is, how do the time-mean changes in
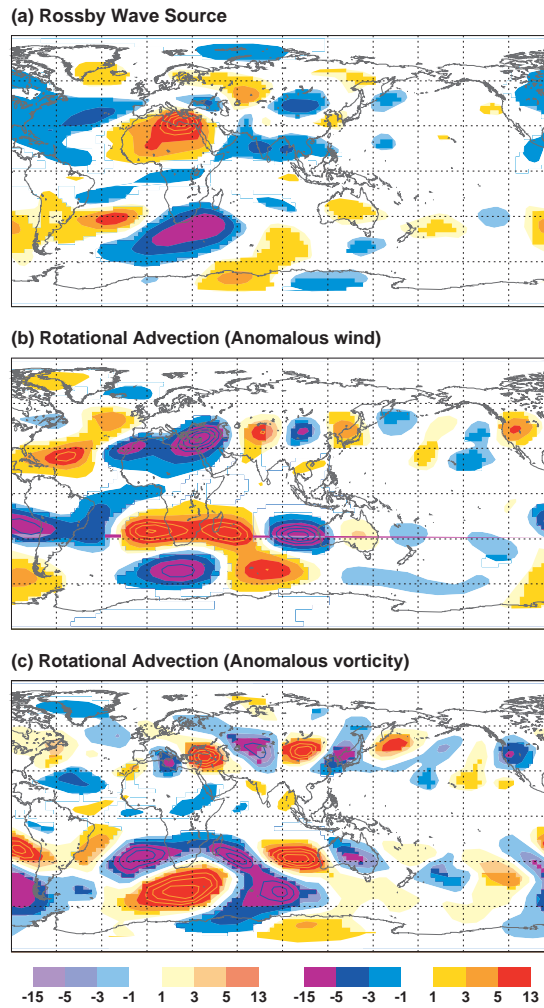
Figure 15: June–August upper-tropospheric vorticity equation response to the change in model aerosol climatology derived from seasonal-mean data from the 40 years of 'seasonal integrations'. (a) The change in Rossby wave source. (b) The change in the advection of absolute vorticity due to changes in rotational wind. (c) The change in the advection by the rotational wind due to changes in the absolute vorticity. All quantities are mass-averaged between 300 and 100 hPa. Shading levels and Contours are displayed at ± 1, ± 3, ± 5,... $\times 10^{-11}s^{-2}$. Features that are statistically significant at the 10% level are shaded using bold colours. Refer to equation (15) for precise definitions of the terms shown.

the terms on the left-hand side of equation (15) balance the Rossby wave source change? Fig. 15(b) and (c) show the first two terms in equation (15), respectively. It can be seen that both these terms are involved in balancing
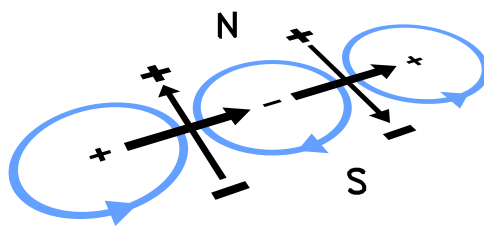

Figure 16: Schematic diagram showing the vorticity advection terms associated with a Rossby wave.

the Rossby wave source anomaly. Away from the strong Rossby wave source anomalies, the two 'rotational advection' terms (Fig. 15c,d) must simply balance each other. This balance can be seen in the extratropics along the waves identified above in the anomalous streamfunction.

Fig. 16 shows a schematic of such a vorticity balance. The blue circles show the location of alternating vorticity anomalies. The thick arrows indicate the mean (westerly) winds that advect these vorticity anomalies downstream. At the same time, the anomalous flow around the vorticity anomalies advects the background planetary vorticity so that the Rossby wave propagates upstream. If the upstream propagation balances the downstream advection then the vorticity anomalies will remain fixed and the Rossby wave will be stationary.

Notice that there is very good correspondence between the upper-tropospheric streamfunction anomalies at 50^oS, south of South Africa in Fig. 14, with the 500 hPa height anomalies in Fig. 2. One can conclude that this extratropical feature is a barotropic stationary Rossby wave response to the aerosol-induced anomalous tropical convection. This feature was the last remaining aspect in Fig. 2 that required explanation. Such a wave explains why a physics change, primarily within the tropics, leads to teleconnections throughout the extratropics. The wave in the Northern Hemisphere is probably 'real', although it is not statistically significant. It is worth mentioning that the waves seen in both hemispheres agree remarkably well with those produced by Ambrizzi et al. (1995) using an idealised barotropic model.

12   The December–February response to the change in aerosol

For completeness, some of results for the aerosol example are briefly repeated for the December–February season. As would be expected, much of the reasoning given for the June–August season carries over to the December–February season. Firstly, we discuss the mean climate for this season, the systematic errors with the old aerosol, and the statistically significant response to the aerosol change.

Figure 17 shows a similar plot to Figure 2 but for the December–February season based on the seasonal integrations started on 1 October for the years 1962–2001. Figure 17(a) shows mean December–February precipitation, low-level (925 hPa) winds and 500 hPa geopotential heights from the observational data. The Southern Hemisphere summer monsoons over South America, Southern Africa and northern Australia together with their associated low-level inflows are clearly evident. In the winter (northern) extratropics, the westerly jet is stronger than it was in the June–August season (Figure 2a) while in the summer (southern) hemisphere, the jet is weaker than it was in the June–August season.

Some of the statistically significant mean errors for the old aerosol integrations (Figure 17b) are reduced when the new aerosol is introduced (Figure 17c). These improvements include a reduction in the erroneous precipitation over the Gulf of Guinea, a beneficial increase in mean precipitation over the equatorial Indian Ocean, a substantial reduction in the extratropical high geopotential height bias over the North Pacific and a reduction in the low geopotential height bias centred over the coast of California. These height biases had been long-standing problems for the ECMWF model (e.g. Jung, 2005). The height changes are reflected in the substantial improvements in mean low-level wind over the North Pacific and (not shown) improvements to synoptic activity in the North Pacific stormtrack region. Interestingly, Miller and Tegen (1998) also found a statistically significant mean response over the North Pacific to the radiative forcing by dust aerosol. In this study, it will be demonstrated that these extratropical anomalies are 'connected' to those in the tropics through the action of upper-tropospheric Rossby waves.

The same physical reasoning as given for the June–August season is used to explain how the local physics in December–February responds to the change in aerosol. Hence the strong reduction in precipitation over the Gulf of Guinea in response to the reduced aerosol (compare Figure 1a,b) is likely to be triggered by reduced short-wave absorption. However, the strength of the change may be particularly strong over the Gulf of Guinea because the reduced short-wave absorption is *not* balanced, over a region of prescribed sea-surface temperature, by surface long-wave, sensible or latent heat fluxes. This explains the great improvement in precipitation in this
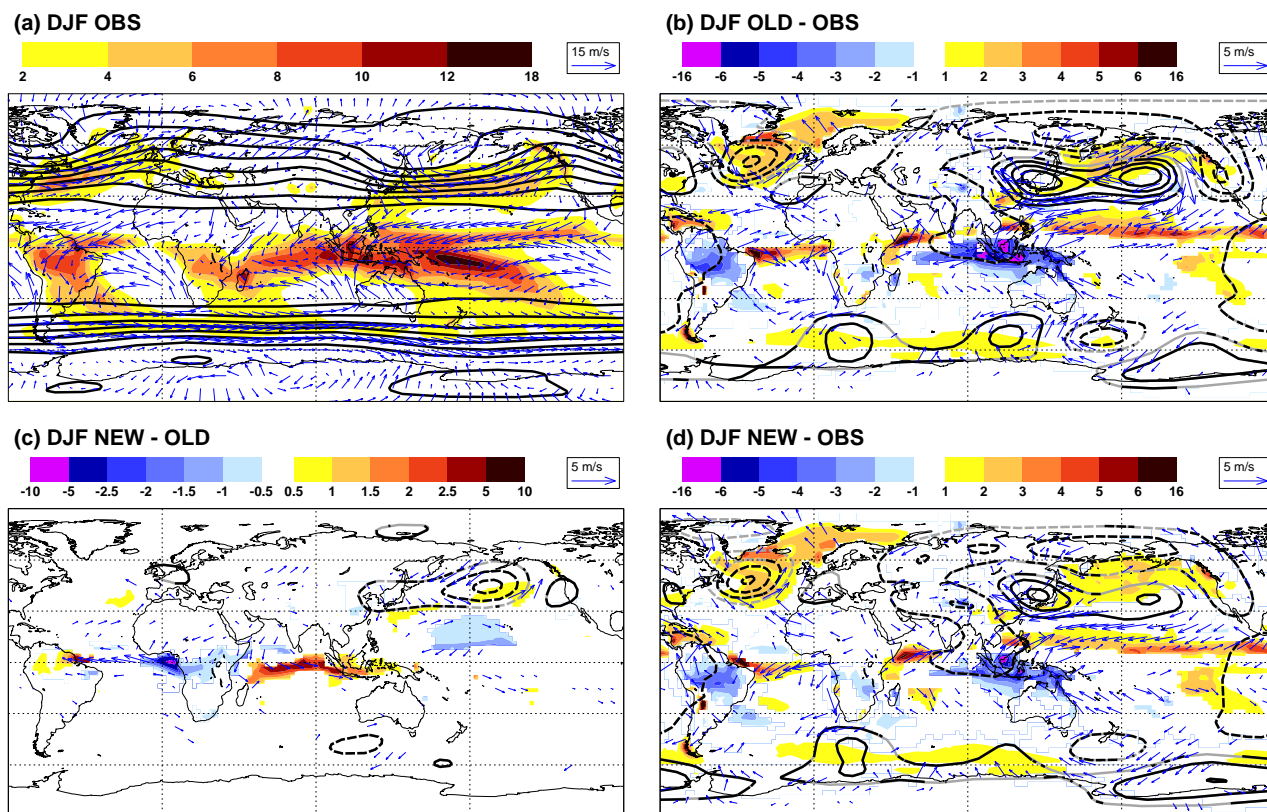
Figure 17: As Figure 2 but for December–February seasonal-means.

region in terms of the atmospheric model being used here but it also highlights a potential difference between atmosphere and coupled ocean-atmosphere simulations. In atmospheric simulations, the ocean effectively has an infinite heat capacity whereas increased downward short-wave could be expected to increase sea-surface temperature in a coupled model and eventually lead to surface fluxes that compensate for the direct atmospheric cooling effect. One may speculate that, in the real world, short timescale fluctuations of aerosol are potentially much more powerful over the ocean than over the land but that this disparity diminishes at longer timescales.

The strong reduction in aerosol over the Sahara (compare Figure 1a and b) also leads to a radiative cooling anomaly (not shown). Although there is no precipitation in this region for this forcing to positively feed-back with at this time of year, the radiative cooling alone is apparently enough to force some descent and upper-level convergence.

With similar reasoning to that for the June–August season, the December–February decrease in precipitation over the Gulf of Guinea and equatorial Africa is likely to force eastward-propagating, upwelling Kelvin waves over the Indian Ocean. The increased precipitation seen over the tropical Indian Ocean in Figure 17(c) is consistent with a diabatic coupling with these waves. Such coupling was speculated for the June–August season above.

Figure 18 shows the same upper-tropospheric diagnostics as in Figure 14, but for the December–February season. The tropical divergent wind anomalies, with anomalous convergence towards the Gulf of Guinea and divergence from the equatorial Indian Ocean are consistent with the latent heating changes associated with the anomalous precipitation. The Rossby-wave source centres, located over Spain/Morocco and over Afghanistan, appear to be related to the tropical response. These Rossby wave source anomalies are consistent with the wave-train in anomalous streamfunction (thick contours) that spreads along the jetstream to the east. Using the same methodology as used for the June–August season, it is again established that this wave is part of a stationary wave solution. The wave clearly connects to the North Pacific anomalies seen in 500hPa geopotential heights (Figure 17c). This stationary wave-train has strong similarities with the circumglobal wave of Branstator (2002).
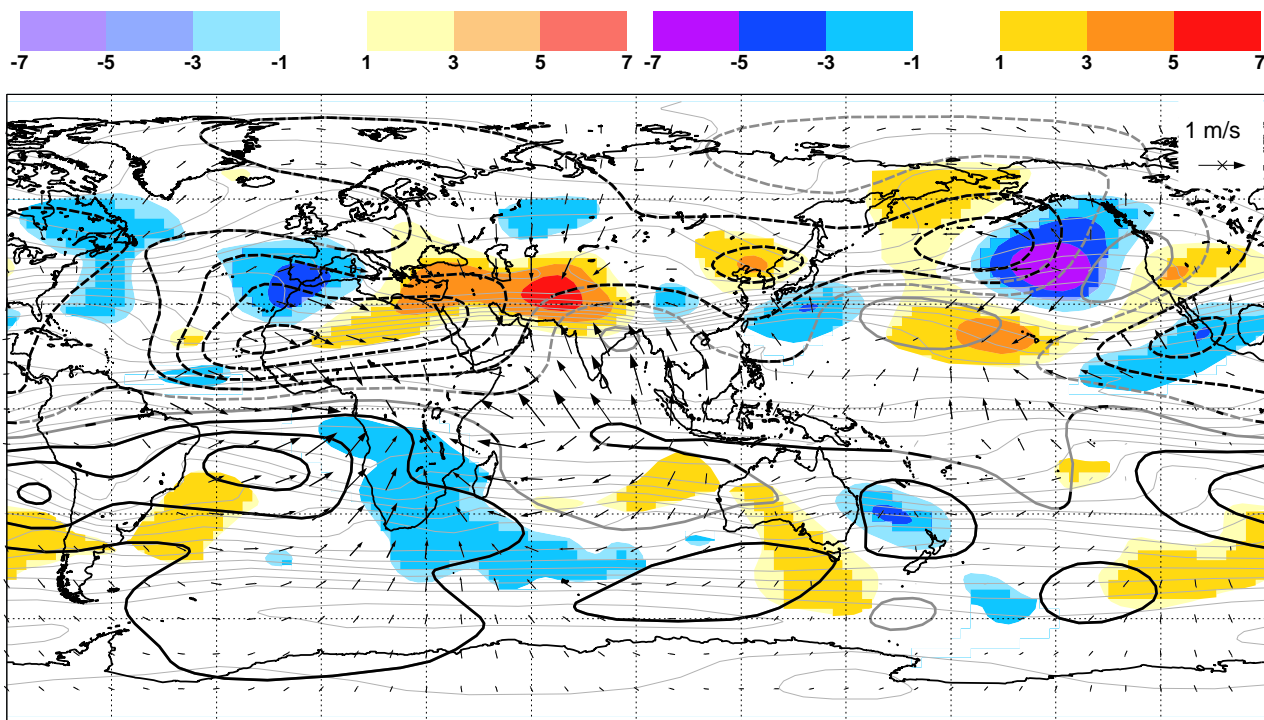
Figure 18: As Figure 14 but for December–February mean change.

The wave's path and zonal wavenumber are highly reminiscent of the Asian jet-stream waveguide highlighted by Hoskins and Ambrizzi (1993). Indeed their barotropic vorticity equation model even predicts a southward turning of the stationary wave as it approaches the west coast of North America just as can be seen in Figure 18.

Interestingly, the amplitude of the wave-train in anomalous streamfunction in Figure 18 is relatively large over the North Pacific. The non-divergent barotropic vorticity equation model of Hoskins and Ambrizzi (1993) also shows that the largest amplitude (in relative vorticity anomaly) for their wave in the Asian waveguide is centred over the west Pacific (at around 140^oE, 30^oN). Although the location is not quite the same as found here, it is clear that non-divergent dynamical processes alone can account for some downstream increase in wave amplitude. In the present study, however, Figure 18 shows that there is also a strong negative 'Rossby-wave source' anomaly centred at 210^oE, 40^oN. The inverted commas are used around 'Rossby-wave source' here because the location is far from the tropics and so this vorticity source cannot be viewed as an independent forcing. Instead, this 'Rossby-wave source' indicates that divergent processes also play a role in defining the stationary wave pattern over the North Pacific. The main component to this 'Rossby-wave source' anomaly is the divergence term rather than the advection by the divergent wind. One possible origin for the negative Rossby-wave source term here is adiabatic vortex-tube shrinkage in the anomalous northward flow. (The anomalous flow is northward because the rotational wind dominates the divergent component shown in Figure 18. Anomalous vortex shrinkage would occur because the isentropic surfaces get closer together towards the pole). However, it is intriguing to note that Figure 17(c) shows increased precipitation in this region. Increased precipitation would be consistent with upper-tropospheric potential vorticity destruction. Other regions in the north Pacific with reduced precipitation (one such centre is visible in Figure 17(c) at around 180^oE, 15^oN) coincide with the two positive Rossby-wave source anomalies. Hence it is possible, if a little speculative at present, that there is some local diabatic modification, or even enhancement, of the tropically-forced stationary Rossby wave taking place over the North Pacific.

13   Conclusions

Physics changes can have global implications. Statistical tests can reveal which aspects are attributable to a given physics change, but to understand the reasons behind the statistically significant response, a set of diagnostic tools is needed. This talk has introduced, from first principles, three methodologies that can be used to help understand the local and global response to a change in (tropical) physics. The methodologies have been applied to a few examples; in particular the example of a recent change in model aerosol climatology at ECMWF.

The 'Initial Tendency' methodology provides a powerful and objective method of assessing errors-in and changes-to the 'fast' physical processes within a model. This methodology has helped in the understanding of the local physics response to the change in model aerosol climatology. Indeed, the approach helps unravel the complex response processes that occur early-on in the forecast and objectively confirms that the new climatology is superior to the old climatology. The initial tendency technique has also been applied to perturbed model ensembles that are used within climate change prediction. In this case, the approach appears to lead to a reduction in our uncertainty in climate sensitivity to carbon dioxide. Through these examples it is hoped that the utility of the Initial Tendency methodology for the development of physical parametrizations has been demonstrated. The Initial Tendency methodology requires that the model comes with its own data assimilation system. Phillips et al. (2004) considered day 5 forecast errors where, arguably, there is less need for a data assimilation system. However, when applied to the aerosol example, day 5 forecast errors would have emphasised the feedbacks with the dynamics rather than the initial physics problems responsible for the forecast error. While the need for the model to come with its own data assimilation system means that the application of Initial Tendencies is presently restricted to numerical weather prediction, it is hoped that results such as those given here will encourage the development of more seamless weather prediction / climate forecasting systems in future.

Equatorial wave theory has been introduced from first principles and shown to be of fundamental importance for understanding the tropic-wide response to a given physics change. This study also highlights the importance, within the real world and in general circulation models, of the coupling between these waves and diabatic processes. Indeed, this coupling can greatly enhance the total tropical response.

To understand the extratropical response, this study has examined stationary-wave vorticity balances within the general circulation model. To build-up to this, the vorticity equation has been introduced from first principles, and its terms explained. Stationary extratropical Rossby-waves are clearly excited by the tropical responses to the aerosol change. These Rossby waves explain the regional extratropical circulation improvements seen when the model aerosol was changed.

Acknowledgements

References

Ambrizzi, T., B. J. Hoskins, and H.-H. Hsu, 1995: Rossby wave propagation and teleconnection patterns in the austral winter. *J. Atmos. Sci.*, **52**, 3661–3672.

Branstator, G., 2002: Circumglobal teleconnections, the jet stream waveguide, and the North Atlantic Oscillation. *J. Climate*, **15**, 1893–1910.

Fouquart, Y. and B. Bonnel, 1980: Computation of the solar heating of the earth's atmosphere: a new parameterization, Beitr. *Phys. Atmosph.*, **53**, 35–62.

Gill, A. E., 1980: Some simple solutions for heat-induced tropical circulation. *Quart. J. Roy. Meteor. Soc.*, **106**(449), 447–462.

Gill, A. E., 1982: *Atmosphere-Ocean Dynamics*. Academic Press. 662 pp.

Greatbatch, R. J. and T. Jung, 2007: Local versus tropical diabatic heating and the winter North Atlantic Oscillation. *J. Climate*, **20**(10), 2058–2075.

Hansen, J., M. Sato, and R. Ruedy, 1997: Radiative forcing and climate response. *J. Geophys. Res.*, **102**(D6), 6831–6864.

Heckley, W. A. and A. E. Gill, 1984: Some simple analytical solutions to the problem of forced equatorial long waves. *Quart. J. Roy. Meteor. Soc.*, **110**(463), 203–217.

Hess, M., P. Koepke, and I. Schult, 1998: Optical properties of aerosols and clouds: The software package OPAC. *Bulletin of the American Meteorological Society*, **79**(5), 831–844.

Horel, J. D. and J. M. Wallace, 1981: Planetary-scale atmospheric phenomena associated with the Southern Oscillation. *Mon. Wea. Rev.*, **109**(4), 813–829.

Hoskins, B. J. and T. Ambrizzi, 1993: Rossby wave propagation on a realistic longitudinally varying flow. *J. Atmos. Sci.*, **50**, 1661–1671.

Hoskins, B. J. and D. J. Karoly, 1981: The steady linear response of a spherical atmosphere to thermal and orographic forcing. *J. Atmos. Sci.*, **38**(6), 1179–1196.

Hoskins, B. J. and M. J. Rodwell, 1995: A model of the Asian summer monsoon. Part I: The global scale. *J. Atmos. Sci.*, **52**(9), 1329–1340.

Hoskins, B. J. and P. D. Sardeshmukh, 1987: A diagnostic study of the dynamics of the northern hemisphere winter of 1985-86. *Quart. J. Roy. Meteor. Soc.*, **113**(477), 759–778.

Jung, T., 2005: Systematic errors of the atmospheric circulation in the ECMWF forecasting system. *Quart. J. Roy. Meteor. Soc.*, **131**, 1045–1073.

Klinker, E. and P. D. Sardeshmukh, 1992: The diagnosis of mechanical dissipation in the atmosphere from large-scale balance requirements. *J. Atmos. Sci.*, **49**, 608–627.

Landshoff, P. and A. Metheral, 1979: *Simple quantum physics*. Cambridge University Press. Pp. 177.

Liebmann, B. and C. A. Smith, 1996: Description of a complete (interpolated) outgoing longwave radiation dataset. *Bull. Amer. Meteor. Soc.*, **77**, 1275–1277.

Matsuno, T., 1966: Quasi-geostrophic motions in the equatorial area. *J. Met. Soc. Japan*, **44**(1), 25–43.

Miller, R. L. and I. Tegen, 1998: Climate response to soil dust aerosols. *J. Climate*, **11**(12), 3247–3267.

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.*, **102**, 16663–16682.

Murphy, J. M., D. M. H. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, M. Collins, and D. A. Stainforth, 2004: Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, **430**(7001), 768–772.

Phillips, T. J., G. L. Potter, D. L. Williamson, R. T. Cederwall, J. S. Boyle, M. Fiorino, J. J. Hnilo, J. G. Olson, S. Xie, and J. J. Yio, 2004: Evaluating parameterizations in general circulation models: Climate simulation meets weather prediction. *Bull. Amer. Meteor. Soc.*, **85**(12), 1903–1915.

Rabier, F., H. Järvinen, E. Klinker, J.-F. Mahfouf, and A. Simmons, 2000: The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. *Quart. J. Roy. Meteor. Soc.*, **126**, 1143–1170.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.*, **108**(D14), 4407.

Reynolds, R., N. Rayner, T. Smith, M. Thomas, D. Stokes, and W. Wang, 2002: An improved in situ satellite sst analysis for climate. *J. Climate*, **15**, 73–87.

Rodwell, M. J. and B. J. Hoskins, 1995: A model of the Asian summer monsoon. Part II: Cross-equatorial flow and PV behavior. *J. Atmos. Sci.*, **52**(9), 1341–1356.

Rodwell, M. J. and B. J. Hoskins, 1996: Monsoons and the dynamics of deserts. *Quart. J. Roy. Meteor. Soc.*, **122**(534), 1385–1404.

Rodwell, M. J. and B. J. Hoskins, 2001: Subtropical anticyclones and summer monsoons. *J. Climate*, **14**(15), 3192–3211.

Rodwell, M. J. and T. Jung, 2008a: The ECMWF 'diagnostic explorer': A web tool to aid forecast system assessment and development. ECMWF Newsletter 117, ECMWF, Shinfield Park, Reading, Berkshire RG2 9AX, UK.

Rodwell, M. J. and T. Jung, 2008b: Understanding the local and global impacts of model physics changes: An aerosol example. *Quart. J. Roy. Meteor. Soc.*, **134**(635), 1479–1497.

Rodwell, M. J. and T. N. Palmer, 2007: Using numerical weather prediction to assess climate models. *Quart. J. Roy. Meteor. Soc.*, **133**(622 A), 129–146.

Sardeshmukh, P. D. and B. J. Hoskins, 1988: The generation of global rotational flow by steady idealized tropical divergence. *J. Atmos. Sci.*, **45**, 1228–1251.

Stainforth, D. A., T. Aina, C. Christensen, M. Collins, N. Faull, D. J. Frame, J. A. Kettleborough, S. Knight, A. Martin, J. M. Murphy, C. Piani, D. Sexton, L. A. Smith, R. A. Spicer, A. J. Thorpe, and M. R. Allen, 2005: Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature*, **433**(7024), 403–406.

Tanre, D., J. F. Geleyn, and J. Slingo, 1984: First results of the introduction of an advanced aerosol-radiation interaction in the ECMWF low resolution global model. In: H. E. Gerber and A. Deepak, eds., *Aerosols and Their Climatic Effects*, pp. 133–177. A. Deepak, Hampton, Va.

Tegen, I., P. Hollrig, M. Chin, I. Fung, D. Jacob, and J. Penner, 1997: Contribution of different aerosol species to the global aerosol extinction optical thickness: Estimates from model results. *Journal of Geophysical Research*, **102**(D20).

Tiedtke, M., 1989: A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Mon. Wea. Rev.*, **117**(8), 1779–1800.

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

von Storch, H. and F. W. Zwiers, 2001: *Statistical Analysis in Climate Research*. Cambridge University Press. 484 pp.

Webster, P. J., 1972: Response of the tropical atmosphere to local, steady forcing. *Mon. Wea. Rev.*, **100**(7), 518–541.

Wheeler, M. and G. Kiladis, 1999: Convectively coupled equatorial waves: Analysis of clouds and temperature in the wavenumber–frequency domain. *J. Atmos. Sci.*, **56**, 374–399.

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

Zender, C. S., R. L. Miller, and I. Tegen, 2004: Quantifying mineral dust mass budgets: terminology, constraints, and current estimates. *Eos Trans. AGU*, **85**(48), 509–512.