Are TOVS temperature retrievals capable of resolving the vertical structure of stratospheric planetary waves?

Nili Harnik[1] and Richard S. Lindzen[2]

Submitted to Annales Geophysicae

December 21, 2001

Contacting addresses:

1. Lamont Doherty Earth Observatory of Columbia University 61 Route 9w
   Palisades, NY, 10964
   email: nili@ldeo.columbia.edu
   phone: 845-365-8493
   fax: 845-365-8736

2. Program in Atmospheres, Oceans, and Climate
   54-1720
   Massachusetts Institute of Technology
   Cambridge, MA, 02139
   email: nilih@tempest.mit.edu
   phone: 617-253-2432
   fax: 617-253-6208

Abstract

This study examines how the main limitations of satellite retrievals, which are coarse vertical resolution and a limited vertical domain, affect the observations of stratospheric planetary wave structure and evolution. The approach is to calculate radiances measured by a virtual idealized satellite equipped with TOVS and sitting at the top of a linear model with stratospheric planetary waves, and invert them to obtain a retrieved field. The retrieved and model waves are then compared. A simplified version of the operational retrieval method is used, where the inverse solution is combined with additional information (a constraint). Many different constraints are used to understand the effect of the additional information, in particular vertical correlations, on the retrievals.

It is found that while TOVS retrievals are capable of resolving the gross amplitude and phase of the waves, as well as their overall vertical propagation direction, they are only marginally able to resolve upward propagating waves from waves that are not purely upward propagating, and are not able to distinguish between the various processes that hinder vertical propagation, like partial reflection from a turning surface, or absorption at a critical surface. The retrievals are also not sensitive to wave damping. Looking at time evolution, the retrievals can capture growth or decay of the waves, and temporary downward reflection, but the detailed onset of these processes is not resolved. The form of the constraint significantly affects the retrievals above 37km, and when the waves have very sharp features, also below that. This may have implications for observations of wave mean flow interactions at a critical surface. The implications for real TOVS observations, both for NESDIS retrievals and directly assimilated radiances, and the relevance to ATOVS, are discussed.

1 Introduction

In the following study we examine how the main limitations of TOVS temperature retrievals affect the observations of stratospheric planetary waves, from the point of view of a stratospheric dynamicist, interested in wave structure and evolution. Planetary waves constitute the largest source of variability in the stratosphere, and are responsible for keeping the zonal mean state away from radiative equilibrium. Resolving their vertical structure is essential for correctly diagnosing many of their dynamical features and their interaction with the mean flow. While the main limitations of satellite retrievals, namely, a coarse vertical resolution and limited vertical coverage, are well known, the way in which these limitations affect the observations of vertical wave structure, is not a priori obvious. A wave propagating on an inhomogeneous medium is likely to have many scales, not all of which are resolved by the retrievals. It is also unclear how the distortions in wave structure will affect our interpretation of wave dynamics. We address these issues by simulating a retrieval system in a model, and explicitly testing how wave structure and evolution are resolved. Since our primary concern is not the details of retrieval technology, but the utilization of retrievals by theoreticians and modelers, we can simplify the problem by using a very idealized retrieval setup.

One of the main uses of the observations is to test the applicability of existing theories. It is essential to understand how the limitations of the observing system are manifested in the observations we use, since some of the discrepancy between theory and observations may be a result of these limitations. A specific example, which partly motivated this work, is determining the extent to which linear wave theory explains observed variability over short periods (Harnik and Lindzen, 2001).

Since linear wave models are sensitive to parameters which are not determined from observations in great accuracy, e.g. details of the basic state or model damping, it is hard to determine, in any given case, whether the observations deviate from modeled waves because of errors in these parameters, because the theory breaks down, or because the observations are inaccurate.

The main source of observations of midlatitude[1] stratospheric planetary waves is temperature soundings from nadir viewing satellites using the Tiros Operational Vertical Sounder (TOVS) instrumentation package. These temperature soundings are currently the only observations of a dynamical property of the stratosphere (as opposed to chemical concentrations) which are continuous in time, span many years, and are global in coverage. We therefore specifically examine TOVS temperature retrievals, and only briefly discuss the relevance to other sounding systems.

Past studies (see Miles and O'Neill, 1989, and references therein) suggest the retrievals can resolve vertical wavelengths of 20km, and there is very little real information in the observational products above 1.5 mb, where the highest weighting function peaks. These studies, however, do not specifically address how these limitations affect observed wave structure and evolution. While there are quite a few studies that test satellite temperature retrievals (see Rodgers, 1976 for references), these studies generally emphasize the ability to retrieve the vertical structure of the temperature field. They do not specifically study the ability to resolve wave structures, which are complex and evolve relatively rapidly. This distinction is important since the temperature varies a lot in the vertical, and most of this variation is in

[1]Equatorial planetary waves are typically of smaller vertical wavelength than midlatitude waves. Most studies of these waves are done by radiosondes, rocketsondes or limb soundings (see Andrews et al. 1987 for a discussion and references).

the zonal mean component, while waves are defined as the deviations from the zonal mean. This means that the variations in the vertical of the total temperature field (roughly 60 K) are larger than typical wave amplitudes (10-20 K), hence it is possible that a retrieval that does quite a good job in capturing the vertical temperature structure essentially captures the zonal mean, but does a poor job in capturing the waves. One notable study, by Graves (1986, see also Karoly and Graves, 1990), tests the retrieval of many dynamical properties, by applying the operational routines used at the time by NESDIS to fields generated by the SKYHI GCM. Some of the fields Graves compares are of wave statistics and wave amplitudes, however, her results do not provide a clear sense of how wave processes are resolved, since she does not study wave structures specifically. Even though the stated resolution of the observations suggests waves are minimally resolved, their time evolution and specific wave processes might get distorted.

In this study we retrieve temperature fields in a simple model that has stratospheric waves, using an idealized observing system. By 'idealized' we mean that the observations are taken simultaneously at all grid points, the instrument weighting functions are assumed to be perfectly known, there are no clouds and no issues of calibration, antenna side lobe corrections, etc. Essentially in this study we determine the information content of TOVS relevant for wave observations, which can be regarded as some form of upper limit on the quality of the observations. We concentrate on testing temperature retrievals, and do not discuss other limitations that enter in deriving geopotential heights and higher order dynamical quantities (e.g. errors in the base level analysis, Karoly and Graves, 1990).

We should note that while TOVS has been collecting data operationally from 1979 to the present, since 1998 a new generation of vertical sounders – ATOVS (Advanced-

TOVS) – has gradually been put to operational use, and will completely replace TOVS in the future. Our study is still very relevant, since TOVS data is used in most existing dynamical studies of stratospheric waves, and will continue to be used in the future (it is the longest data set available and a lot of experience has been accumulated using it). In addition, even though ATOVS has different weighting functions than TOVS, the results of this study are still useful for interpreting ATOVS data since the two instrumentation packages are based on the same principals and their limitations are qualitatively the same, though there are obviously quantitative differences.

We start in section 2 by describing the setup of the numerical model-retrieval system, and discuss the main characteristics of the retrieval of a single profile (section 2.3). In section 3 we study the retrieval of various wave fields, representing different wave processes. We summarize and discuss our results in section 4. The calculation of model generated temperature fields, and TOVS radiances, and the minimum variance retrieval method that we use are described in the appendices.

2 The virtual satellite: experiment setup

Our interest is in understanding the main implications of the inherent limitations of the nadir soundings when applied to the observations of stratospheric waves, rather than in testing out a specific retrieval method or instrument. For this, an idealized approach suffices, which allows us to simplify our calculation considerably.

We use a linear quasi geostrophic (QG) model to obtain temperature fields containing stratospheric waves. We specify a vertical profile of Brunt Vaisala frequency, a zonal mean wind as a basic state and a geopotential height perturbation at the bottom of our model as forcing. The model is described in Harnik and Lindzen (2001),

and details which are specific to this study are described in appendix A.

We assume a satellite with TOVS instruments is sitting above our model. Using TOVS weighting functions we calculate the radiances it would measure (*the forward problem*) and apply some inverse technique to retrieve the temperature fields from these radiances (*the inverse problem*).

We repeat the forward-inverse calculation at each model grid point to obtain a three dimensional retrieved temperature field. To obtain the retrieved *wave* field, we subtract the zonal mean. Operationally, the raw data from TOVS undergoes a series of corrections (e.g. limb and antenna side-lobe corrections, cloud contamination and extrapolations due to differences in the scan pattern of the three instruments) to produce the clear-column radiances used for the retrievals (see Smith et al. 1979, and Kidwell, 1986, for details). We avoid these issues by assuming clear skies and that the satellite samples all grid points simultaneously and directly from above. Another problem we avoid is aliasing related to asynoptic sampling (Salby, 1982)[2].

[2]Unlike a radiosonde network, where all measurements are taken simultaneously at specified times, a satellite samples the domain in a continuous scan, and all the measurements taken over one day are combined to a single map. Salby (1982) showed that this results in an aliasing of zonally propagating perturbations with periods shorter than 1-2 days. Periods of observed planetary Rossby waves are typically much longer, but observed structure changes that involve a tilting of wave phase lines on daily time scales (Harnik and Lindzen, 2001) may be aliased since they appear as a zonal propagation at some levels. We test this by asynoptically sampling an analytically specified wave undergoing structure changes as observed. The main effect is to decrease wave amplitudes. For realistic structure changes we find a noticeable decrease only in the upper stratosphere (where the observations are least reliable anyway), and the fastest changes observed would result in a decrease of less than 20% in amplitude.

2.1 The forward problem

The forward problem consists of calculating the radiances from the model temperature field as follows. Given a temperature profile ($T(z)$) and an instrument channel response frequency ν, the total radiance viewed from a satellite looking directly down is:

$$I_\nu(\infty) = I_\nu(z_s)\mathcal{R}_\nu(z_s) + \int_{z_s}^{\infty} B_\nu(T(z))\frac{\partial \mathcal{R}_\nu}{\partial z}dz \tag{1}$$

where z is log-pressure, and z_s is the lowest model level, which is chosen at 14km, and $B_\nu(T(z))$, and $\mathcal{R}_\nu(z)$ are the monochromatic Planck function and transmittance function, respectively[3]. Definitions of the variables and a derivation can be found in atmospheric radiation or remote sensing text books (e.g. Liou, 1980; Stephens, 1994; Kidder and Vonder Haar, 1995). The bottom boundary condition, which we apply at z_s, is described in appendix B.

The vertical derivative of $\mathcal{R}_\nu$ is generally referred to as the *weighting function*: $W_\nu(z) \equiv \frac{\partial \mathcal{R}_\nu}{\partial z}$. The shape of the weighting function indicates the levels that contribute most to the radiance at the given frequency ν. The weighting functions depend on the concentration of the corresponding emitter (CO_2 or O_2 for TOVS), humidity and temperature, and uncertainties in these profiles introduce an uncertainty into the weighting functions. For simplicity, we assume the weighting functions are perfectly known.

[3]Since an instrument channel is sensitive to a narrow band of frequencies centered around the frequency ν, the radiative transfer equation has to be averaged over the instrument response function. We assume equation 1 has already incorporated this averaging, and the transmittance and weighting function profiles include this.

Figure 1 (see also table 1) shows the TOVS weighting functions that peak at or above 14 km, which is the lower boundary of our model. See appendix B for references and how we calculate the weighting functions. The solid lines are the High Resolution IR Spectrometer (HIRS) and the Stratospheric Sounding Unit (SSU) channels. The dashed line corresponds to the Microwave Sounding Unit (MSU) channel 4, which we do not use in our study. Dropping MSU 4 does not affect the main results of our study, since it peaks in a region that is well covered by other channels, adding little new information to the retrievals. Operationally, one of the main purposes of MSU 4 is to provide a way of correcting measurements from other channels for cloud contamination (which is not a problem in the current idealized treatment).

We assume that all channels have the same frequency, corresponding to the SSU wavenumber of 668 cm^{-1}. This simplifies the inverse problem by making it linear because we can solve for $B(z)$ instead of for $T(z)$. For reference, the true channel wavenumbers are listed in table 1. We use a nonlinear Chahine method (see section 2.2) to compare the retrieval with a single frequency to the retrieval with actual frequencies, and find very little difference. It is, of course, essential to use the same frequency in the forward and inverse problems. Such linearization is much harder to justify with MSU 4 since its microwave frequency is significantly different from all other IR channels, which is another reason we drop it.

A general characteristic of inverse solutions is a trade off between resolution and sensitivity to noise (e.g. Mateer, 1965, Rodgers, 1990). It is therefore important for us to test the sensitivity of the retrievals to instrument noise. We do this by adding noise to the radiances that are calculated from equation 1 before inverting them to get temperatures. The noise we use is described in appendix B.

2.2 The Inverse problem - general setup

The inverse problem consists of solving a set of equations 1 for the different instrument channels to get a vertical profile of temperature. The problem is inherently ill posed, since we are trying to solve for a continuous profile using a discrete set of measurements. Physically, the satellite senses radiances emitted from a layer, hence we can only obtain a temperature profile with low vertical resolution. There is a great body of literature that deals with the inversion of satellite IR measurements to obtain temperatures (see references in Rodgers' 1976 excellent review paper). Different approaches include solving for the temperature at M discrete levels (Chahine, 1970), solving for the expansion coefficients of the vertical temperature profile around a set of M basis functions (Mateer, 1965), and using statistical regression onto an existing climatological data set (Smith and Woolf, 1976). The approach that is currently used to produce operational temperature retrievals at NESDIS is a Minimum Variance (MV) method (Rodgers, 1976). In this method the observations are combined with some additional information, typically a climatology, in order to make the problem well-posed. In a MV retrieval, two estimated profiles of temperature are combined, one based on the actual measurements, referred to as the *inverse solution*, and one based on some additional information that we have, referred to as the *constraint* profile. The *retrieved* profile is a weighted average of the these two estimated profiles, weighted by their inverse covariances. A more recent variation of the Minimum Variance method is to combine the inverse solution with a numerical weather prediction model forecast. This is currently done operationally at ECMWF (Eyre et al. 1993; Andersson et al. 1994) and NCEP (Derber and Wu, 1998).

In the current study, we retrieve temperatures using a linear version of the MV method, with two kinds of constraints. The first is a diagonal constraint (a diagonal

error covariance matrix), which means all the vertical correlations in the retrieval are due to the observing system. The second is a non-diagonal constraint (a non-diagonal error covariance matrix) meaning the constraint also contributes to the vertical correlations of the retrieval. The retrieval method and the constraints are briefly described in appendix C. In the operational retrievals the constraint has vertical correlations.

What allows us to use the simple MV retrieval scheme to determine the basic limitations of real observations is the following. The inverse problem consists of optimally combining the radiance measurements with some additional information that is necessary to make the inverse problem well behaved. One of the main points of Rodgers' (1976) review is that different retrieval schemes differ in the kind of additional information that is used, and unless this additional information contains additional observations taken at the same time from another source, the true vertical resolution of the observations is the same for all methods. This is also true for direct assimilation of radiances into a numerical weather prediction model since TOVS radiances are (as far as we know) the only observations that constrain temperatures in the middle and upper stratosphere that go into the assimilation products. To determine the effects of the additional information, we compare retrievals with different kinds of constraint. We also compare the MV retrieval to a Chahine (1968, 1970) method, which solves for the temperature at six discrete levels[4], using a nonlinear iterative scheme. Since Chahine's method does not use additional statistical information, this comparison highlights the robust features from those that depend on the specifics of the statistical constraint.

[4]Chahine bases his algorithm on the fact that the weighting functions have a well defined peak. Since HIRS channel 1 has a very wide weighting function (see figure 1), it degrades the retrieval substantially and we need to drop it. As a result, we retrieve temperature only on six levels.

2.3 Characteristics of the retrieval of a single profile

To understand the limitations of the retrievals in resolving wave structures, it is essential to first understand how the observations and the constraint are combined in the retrieval of a single profile. Even though similar calculations have been done in the past (e.g. Rodgers, 1976, 1990), it is useful to present them using our specific idealized setup.

We start by understanding the effect of a diagonal constraint. For simplicity we use a constant variance with height. Figure 2 shows a temperature profile (line with solid circles), and the various diagonal MV retrievals, all using the same constraint profile (thin dotted line) but different values of variance. Also shown are the true minus retrieved profiles for each of the retrievals, to highlight the deviations. For comparison we also show the Chahine retrieval (circles), which points out from the start the inherent limitations of the satellite observations. Even if we were to do a perfect job by retrieving the exact temperature at the six levels, we are retrieving only six points. There is no information above 1.5 mb and the resolution is at best as good as the distance between the peaks of the weighting functions.

The general behavior of the retrieval is to follow the inverse solution (i.e. the true profile) at the lower levels, where the variance in the measurement is smaller than the variance in the constraint, and to follow the constraint at higher levels, where there are no observations. In between, the solution is a combination of the two, where we find that a larger constraint variance allows the retrieval to follow the inverse solution more closely and over a larger region[5]. As a result, the retrieval solutions are best below 30-35 km, where they are within 1-2 K of the true temperature. Above that,

[5]Note that this behavior is specific to the diagonal retrieval, where the constraint errors are uncorrelated in the vertical.

the errors become larger, and dependent on the difference between the constraint and true temperature profiles.

The decrease of retrieval errors with variance is only true when we use the exact radiances. Since in reality we always have errors, it is important for the retrievals to have a low sensitivity to noise. We calculate the sensitivity to noise by comparing the retrievals using noisy and exact radiances (see section 2.1 and table 1). The resultant temperature error statistics are relatively constant with height[6], and as expected, larger values of constraint variance are associated with a larger sensitivity to noise. The standard deviations are 1, 1.5, and 2.5 K for a variance of 5, 10, and 20 K, respectively. Locally, errors can reach values as high as 4, 7, and 15 K for these variances, but these are generally isolated points. The general noise levels correspond more to the RMS values.

Mathematically it can be shown that the retrieval takes the largest vertical structures from the measurements (the inverse solution) and small vertical scales from the constraint, while in between scales are taken from both (Rodgers, 1990). This is done by projecting the retrieval, the inverse solution, and the constraint onto the eigenvectors of the Averaging Kernel Matrix $\mathbf{A_{km}}$ (defined in appendix C). The eigenvalues are the partial contribution of the inverse solution, and the largest eigenvalues correspond to the largest vertical structures (meaning the inverse solution mostly determines them), while the smaller eigenvalues correspond to smaller vertical scales (the constraint determines them). Figures 3 and 4 show the largest eigenvalues and eigenvectors of $\mathbf{A_{km}}$ for a few values of constraint variance. We see that the ver-

[6]This is a result of the error covariance matrix being diagonal, with a constant value of variance. Also, errors are larger (almost by a factor of two) near the bottom boundary, most probably because we do not have a troposphere in our retrievals.

tical scales of the eigenvectors decrease with decreasing eigenvalue, and that as the variance of the constraint increases, the eigenvalues of $\mathbf{A_{km}}$ increase, meaning more structures are determined by the inverse solution. Optimally, the retrieval should draw from the inverse solution as much information as it contains, and no more. A crude estimate of the number of independent structures that are resolved by TOVS is 3-4, which is obtained from the weighting functions themselves (by looking at the eigenvalues of $\mathbf{KK^T}$, Rodgers 1976, section 4a; see also Mateer, 1965). Looking at figure 3, this corresponds to an optimal constraint variance of about $7 - 10K$, which has RMS errors of roughly 2 K below 35 km. Note that the fifth eigenvector of $\mathbf{A_{km}}$ (figure 3) is the most stable structure in the sense that it is the least affected by the variance of the solution. This suggests it is some measure of the inherent resolution of the observing system. A similar structure shows up as the characteristic scale of errors when the solution is weakly constrained (see results in the next sections). These correspond to a vertical wavelength of roughly 20km.

The effect of vertical correlations in the constraint is best illustrated by looking at the response functions (appendix C), which are the response of the observing system to a spike of temperature perturbation at a given level. The nondiagonal constraint we use (described in appendix C) is specified from a climatology that we create from a time dependent version of our model, where we superpose a transient and a stationary wave number 1. The resultant waves have a vertical structure which is different from the waves we retrieve. Figure 5 shows the response functions of the non-diagonal constraint retrieval, and a corresponding diagonal retrieval where only the diagonal elements of the error covariance matrix are retained, while the rest are set to zero[7].

[7]The difference between the current diagonal covariance matrix and the ones tested in the previous section is that the variance varies with height, whereas the constraint of the previous section assumed

The values of these diagonal terms are shown in table 2.

We see that the main effect of the vertical correlations is to spread the response to non-adjacent layers. The response functions have two or three peaks, instead of a single major one. This remote response is the result of having waves in the climatology from which the constraint was calculated. The largest response is at high levels, in some cases, exceeding the local response, This is a result of the increase of wave amplitude with height in the model run used for the constraint calculation. Note however, that the remote responses to perturbations at various heights may cancel each other.

We also test the sensitivity to instrument noise, as we do with the diagonal constraint, and we find that the main effect of having vertical correlations is to cause the response to errors in the radiances to increase at upper levels. The RMS temperature error is roughly 0.5-1 K below the second highest weighting function (5mb), it doubles up to the top weighting function (1.5mb), and increases to four times as much above that.

3 Results - retrieving various wave structures

In this section, we present the results of retrieving various temperature fields from our QG model. We generate waves on various kinds of basic states, using various damping fields, in order to see how the observations resolve different wave processes. We use a stationary perturbation of zonal wavenumber 1, but our main results extend to other zonal wavenumbers and phase speeds. See appendix A for details about the model and how we calculate the true temperature fields. In all figures latitude is

a constant variance. The current variance is shown in table 2.

nondimensionalized by the radius of deformation ($1272.5km$), and the vertical grid below 49 km is at the peaks of the weighting functions.

We start with the simplest case, of a wave propagating upward through the stratosphere, and dissipating only in the sponge layer in the mesosphere. The solid line in figure 6b shows the damping rate below 63 km (our model extends beyond the domain shown in the figures). Figure 6a shows the amplitude and phase of the wave, while figure 7a shows a longitude-height section at latitude $y = 2.3$. The resulting wave has a very simple vertical structure, as expected from linear wave propagation theory (Charney and Drazin, 1961), with an amplitude increasing with height due to the density effect, and a relatively constant westward phase tilt with height. We retrieve this wave field using a diagonal MV retrieval with a constant variance of 7°K (figures 6c and 7d), and a nondiagonal MV retrieval using the constraint of figure 5 (figures 6e and 7g). We also show the true minus retrieved wave amplitudes, for the diagonal (6d) and nondiagonal (6f) retrievals. We see that both retrievals do quite a good job in capturing the amplitude and phase structure of the wave below 45 km (the peak of the highest weighting function). This is not surprising since the vertical wavelength is much larger than 20km, and the vertical structure of the wave is simple. The amplitude errors below 45km are roughly 2°K for the diagonal retrieval and less than 1°K for the nondiagonal one. In this case, the nondiagonal constraint, which contains basic wave structure information like an exponential increase of amplitude with height and a westward phase tilt with height, improves the retrieval compared with the diagonal case.

Above 45km, the retrievals tend to follow the constraint. As a result, errors become large, and the two retrieval methods behave very differently. The diagonal retrieval wave amplitude tends to zero with height, and the phase tilt is smaller than

for the true waves (resulting in phase errors that increase with height). This is because the constraint contains no waves. The phase structure of the solution is biased towards the structure at the highest level at which the true waves are felt, roughly the peak of the top weighting function. The nondiagonal retrieval, on the other hand, does have waves in the constraint. Correspondingly, the amplitude continues to increase with height above the peak of the top weighting function, and the phase continues to tilt westward with height, resulting, *in this particular case*, in a better retrieval.

The effect of adding random noise to the radiances (not shown) is to add a white noise field to the temperature of roughly $2 - 3K$. The resultant retrieved wave field looks a bit noisy, but the overall large scale pattern is still evident. When a Fourier decomposition is made and only wave one retained, there is a very small difference between the exact and noisy fields, meaning the noise does not project significantly onto wave 1.

The wave shown above is very simple in the sense that the density effect causes amplitude to increase with height, but otherwise, the structure of the wave does not vary with height. In the real stratosphere, we expect waves to have a more complex vertical structure, because processes like damping, reflection, absorption, and transience, cause the wave structure to change with height. We show a few examples of such cases, which clearly demonstrate the limitations of the retrievals.

The first process we examine is damping. Estimates suggest that thermal damping peaks in the upper stratosphere with damping time scales reaching values of about 2 days at 50 km (Dickinson, 1969). There is also momentum damping, due to wave breaking (both gravity waves and Rossby waves). In our model we crudely parameterize this effect as a linear Rayleigh damping. From models, we find that wave structure is sensitive to the exact form and strength of damping, but in general, damping causes

the amplitude of the waves to decrease with height in damping regions. When relevant, damping also results in a more westward phase tilt with height, because it reduces the relative amount of downward reflected waves in the total wave field. Here we show an example where we add damping in the upper stratosphere, to the upward propagating wave shown above, and compare the retrieval of both cases. We use equal damping rates of momentum and temperature, which are shown in figure 6b (dashed line). Repeating the same calculations but with thermal damping only (using the damping rates of Dickinson, 1969) gives similar results. We refer to the simple wave of figures 6 and 7a as the 'undamped' wave (even though it is damped in the mesospheric sponge layer). Figure 7b shows a longitude-height section of the damped wave, shown at the same latitude as the undamped wave. Also shown is the difference between the damped and undamped waves (7c). The difference below 37 km is small, and is a quarter wavelength out of phase with the actual wave fields, meaning the main effect of damping is a slight shift in phase. Above 37 km, the damped wave amplitude is smaller than the undamped one, with the amplitude actually decreasing with height above 50 km, reaching a magnitude of 6°K where the undamped wave is greater than 39°K.

The two retrieved fields, along with the corresponding undamped minus damped wave fields are also shown in figure 7. The most striking result is that the difference between the undamped and damped waves is much smaller in the retrieved fields than in the true ones. While the diagonal retrievals of both have characteristics of a damped wave, with the amplitude decreasing in the upper stratosphere, the nondiagonal retrievals both look like an undamped wave, with wave amplitude increasing with height throughout the stratosphere and beyond. There are some differences between the damped and undamped waves in the retrievals, but unlike the true waves,

these differences are quantitative, not qualitative. Given that we expect damping to become large only in the upper stratosphere, we conclude that the observations can not tell us much about the damping that the waves feel.

Two other phenomena that affect vertical wave structure are absorption at a horizontal critical surface (the zero wind line for stationary waves), and downward reflection. Linear waves propagating to a critical surface will get absorbed (Dickinson, 1968)[8]. This will cause an abrupt termination of the wave geopotential height, and a corresponding sharp (in the vertical) temperature perturbation near the critical surface. Horizontal critical surfaces form in the stratosphere every spring in both hemispheres, when the polar vortex breaks down, and in mid winter in the Northern Hemisphere during sudden warmings. Throughout the summer, there is a critical surface in the lower stratosphere. Downward reflection also affects wave structures. Recently, Harnik and Lindzen (2001) developed a diagnostic of reflecting surfaces, and showed that such surfaces occasionally form in the upper stratosphere during winter. The waves respond by reflecting downwards, which affects their vertical structure. Downward reflection causes the vertical phase tilt with height to decrease, and the amplitude to decrease above the reflecting surface. When there are two reflecting surfaces, with an evanescent region in between, we have partial reflection and partial transmission, which affects the vertical structure of the wave in a manner which depends on the specifics of the basic state and the wave. For the purpose of this study, We constructed two basic states that are qualitatively similar in the lower stratosphere but one has a critical surface at 50 km, at which the waves get absorbed,

[8]Nonlinear waves can be reflected from the critical surface, rather than be absorbed in it, but in both cases we expect a sharp termination of the perturbation at the critical surface, and a corresponding sharp temperature structure.

and the other has two reflecting surfaces, at 37 and 45 km, with an evanescent region in between, resulting in partial reflection and partial transmission. The zonal mean wind of both states is shown in the top plots of figure 8. We have a small constant damping on temperature and momentum (time scale of 25 days), needed to get rid of the singularity at the critical surface, and a mesospheric sponge layer. The critical surface is marked by a thick line (top right), and the evanescent region, calculated as in Harnik and Lindzen (2001), is shaded (top left). The amplitude and phase of the steady state wave 1 temperature perturbations are shown in the bottom plots of figure 8 and a longitude-height section of both waves at $y = 2.3$, and their difference (partially reflected minus absorbed waves) are shown in figure 9. We see that the two waves have a similar structure in the lower and middle stratosphere, below the lower reflecting surface. At higher levels, one wave gets absorbed at the critical surface, resulting in a very concentrated temperature peak there, and no perturbation above. The partially reflected wave, on the other hand, has vertical phase lines in the evanescent region, and the characteristics of an upward propagating wave above the second reflecting surface, with an increasing amplitude and westward phase tilt with height.

Longitude-height sections of the retrieved wave fields and the corresponding difference fields are shown in figure 9 (middle and bottom rows).[9] There are a few points to note. First, the retrievals are not capable of distinguishing between the partially reflected and absorbed waves. The two waves look qualitatively similar in each of the retrievals. The quantitative differences (shown on the right) are noticeable, but are much smaller than for the true waves. The two retrievals differ greatly from each other above 37km, with the diagonal (nondiagonal) retrieval waves looking like

[9]For this run we show the 10°K variance diagonal retrieval, but 7°K gives similar results.

damped (propagating) waves in the upper stratosphere. The main characteristics of partial reflection, namely, the vertical phase lines in the evanescent region with propagation above, are not captured by the retrievals. Likewise, the main characteristics of absorption, which are the termination of the perturbation with a sharp peak at the critical surface, are absent from the retrieved fields. Below 37 km, the overall phase and amplitude of the waves, are captured by the retrievals, but there are obvious distortions. Compared to the simple propagating wave of figure 6, the retrievals do not do as well. This is because downward reflection and absorption introduce smaller scales. With the exception of the diagonal retrieval of the absorbed wave, the retrievals smooth out the true wave peak at 27 km, and move it down. Also, the change in phase structure at 29.6 km (for both waves) occurs at 37 km in the nondiagonal retrieval. Comparing to the simple wave of figure 7(a,d,g), the retrievals do capture some of the qualitative differences between the pure upward propagation and the other scenarios, but these differences are much more pronounced in the true fields.

One of the characteristics of the nondiagonal retrieval is that it introduces a propagating wave structure above 37km, regardless of whether there is a true perturbation there or not. We now ask whether this is a feature of the nondiagonal retrieval only in the upper stratosphere where the real information content of the retrievals is low, or whether spurious waves can be introduced even lower down. To test this we run our model with stationary wave 1 forcing, and a basic state that is characteristic of early summer, with a critical surface at around 20km, and easterly winds above that. Figure 10 shows the wave 1 temperature amplitude and a longitude height cross section, at latitude y=2.83. As with the upper stratospheric critical surface, the wave is absorbed at the critical level, and the corresponding temperature wave field is cut

off sharply. Also shown in figure 10 is the non-diagonal MV retrieval of these fields, using the constraint of figure 5. The vertical correlations of the retrieval introduce waves into the upper part of the domain, even in regions which the satellite can see. The retrieved field is essentially a superposition of the response functions (figure 5) to the perturbations at all levels. Since there is cancellation between the response to different levels (compare first and third plots in figure 5), a sharp feature is more likely to cause a large spurious wave. Note that the retrieved wave is qualitatively like the first plot of figure 5 which shows the response of the MV retrieval to a delta function of temperature at 19km. Also, the vertical structure of the retrieved wave corresponds quite well to the fourth eigenvector of the Averaging Kernel Matrix, which has an eigenvalue of 0.5, meaning the constraint contributes as much as the inverse solution. Since this vertical scale is larger than the vertical resolution of the operational product (which has nine levels in the stratosphere), we are faced with the problem that the wave does not look obviously spurious. When we repeat the retrieval using a diagonal error covariance matrix with a constant variance of 10 K, the retrieval does a much better job at higher levels, where it does not have waves. Unlike previous cases, where below 37 km the different retrievals did qualitatively the same, the form of the constraint matters a great deal in this case, even in the lower and mid-stratosphere. We discuss the implications for operational retrievals later on.

Finally, all the cases discussed so far were of stationary waves. Transience, however, can introduce smaller vertical scales to the waves, with one obvious scale being that of a wave front. Stratospheric waves are observed to vary considerably in time (e.g. Hirota and Sato, 1969), as a result of variations in their tropospheric source and the basic state. While much of this evidence comes from the same satellite observations we are testing, few will argue that the variability itself is in question. Radiosonde

based observations of waves in the upper troposphere and lower stratosphere, and the fact that the observed time evolution is continuous and seems physical, both indicate that the variability is real. Here we test how the retrieval process might distort time evolution. We examine the retrieval of various time evolving waves, generated by a time dependent version of our β-plane model, in which we specify a time dependent basic state and forcing. The results we show are from a run shown in Harnik and Lindzen (2001) where an upper stratospheric reflecting surface shifts downward from 50 to 40 km (see figure 12 of Harnik and Lindzen) over a period of 2 days, and the source of the wave decreases at the same time, causing the wave to temporarily reflect downward. This run was done in order to explain observed changes in wave structure in late southern hemisphere winter. Figure 11 shows longitude-time sections of the temperature field at $y = 2.3$, for days 6, 7, and 8 of the model run (top row), along with the nondiagonal retrieval of these wave fields (bottom row). The reflecting surface moves down between days 5-7. In response, downward reflection starts in the upper stratosphere on day 7, and spreads downwards (day 8). The nondiagonal retrieval captures the downward reflection (phase lines tilt eastward with height on day 8) but not the details of its development; there is no evidence of downward reflection on day 7, while on day 8, reflection has spread all the way down, whereas in the true wave the downward reflection front only reached 24 km. These results imply that while the retrievals are capable of resolving large changes in the amplitude and propagation direction of the waves, they are not able to resolve the details of how these changes actually evolve in time.

4 Summary and discussion

In this study we examine the ability of TOVS measurements to observe the vertical structure and evolution of midlatitude stratospheric planetary waves. Essentially, we determine what are the implications of the limitations of the retrieval system, which are low vertical resolution and diminishing observations in the upper stratosphere, for resolving the spatial and temporal variations in stratospheric wave structure. Our results are first of all useful for interpreting studies based on TOVS temperature retrievals, which are the only continuous long term global observations of a dynamical quantity of the stratosphere. These results are also useful in pointing out potential problems of other data sets, including assimilation products that assimilate radiances directly, and retrievals based the newer nadir sounding system ATOVS.

We find that while the retrievals are capable of observing the gross amplitude and phase structure of the waves, they are not capable of observing important details. The retrievals can tell us the overall amplitude of the wave (with some vertical averaging), the longitudinal position of the ridges and troughs, and whether the wave is, on the whole, propagating up or down. This means the retrievals can distinguish between an upward propagating wave and a *fully* downward reflected wave. Since there is damping in the upper stratosphere, full downward reflection is only observed as a temporary response. The retrievals can marginally distinguish between pure upward propagation and 'hindered' propagation (partial reflection or absorption at a critical surface), but they can't distinguish between the different 'hindering' mechanisms. We may be able to determine that a certain process occurs from looking at the basic state (for example, if it contains a critical surface or a reflecting surface). The retrievals are also not sensitive to the effects of damping. This means wave observations cannot

be used to study damping processes. The time evolution of the waves is also affected by the retrievals. The retrievals are able to see large overall increases or decreases in wave amplitude, as well as a temporary full downward reflection, but the details of how these changes occur, for example, the propagation of a reflection front, are not resolved.

Our study also indicates that observations based on TOVS retrievals are limited for the purpose of studying wave-mean flow interactions at a critical level, because of the sharp vertical structures involved. This is especially relevant for the breakup of the polar vortex at the end of winter and for sudden warmings in the northern hemisphere winter. The retrievals are unable to resolve the sharp features, and may also introduce spurious waves to the retrieval through the vertical correlations in the constraint. A diagonal constraint, on the other hand, will not introduce spurious waves above the critical surface. This highlights the fact that the form of the constraint matters when sharp features exist, because they project onto smaller scales which are taken mostly from the constraint. The constraint also matters for the retrievals in the upper stratosphere. It is therefore important to understand how the operational retrievals are done.

The data that is most commonly used for stratospheric dynamics studies is based on NESDIS temperature retrievals. NESDIS uses a nonlinear MV retrieval (equation 99 of Rodgers, 1976; Smith et al. 1979), with a non diagonal constraint that is calculated from a dynamic set of radiosonde profiles measured in the two weeks prior to the retrieval. Since radiosonde data does not extend into the upper stratosphere (it is also very sparse in the southern hemisphere), the radiosonde profiles are extrapolated upwards using the statistics of rocketsonde data. The rocketsonde data set is divided for this purpose into three latitude bands (high, middle and low), and into seasons.

As a result, we expect more waves in the error covariance matrix of the constraint in winter compared to summer, and in active periods compared to quiet ones.

A more recent approach to temperature retrievals is to assimilate radiances directly into a numerical forecast model (e.g. Eyre et al. 1993, and Andersson et al. 1994, for ECMWF; Derber and Wu, 1998, for NCEP). This essentially means that the inverse solution, namely the measured radiances, is combined with a model forecast to obtain the analyzed temperatures, which are the retrieval. In terms of forecasting, this is a significant improvement over assimilating temperatures because only actual observations are assimilated, as opposed to also assimilating the climatological constraint. In terms of stratospheric temperature retrievals, the real information content is the same for both products, since radiosondes hardly reach the middle and upper stratosphere, and TOVS is the only source of observations there[10]. The main difference is that the assimilation product fills in the information on small vertical scales and extrapolates the observations to the upper stratosphere based on a model run, rather than on vertical correlations from a climatology[11]. Note that the phase tilt of the waves, as well as their amplitude, can be affected by the form of model damping (e.g. Harnik and Lindzen, 2001), which is one of the most uncertain parameterizations in stratospheric models.

In the process of assimilating radiances with the model forecast, an error covariance matrix has to be specified (Derber and Wu, 1998). This matrix is generally constant with time and is calculated from long model integration statistics. Vertical

[10]Another source of observations in the middle and upper stratosphere is satellite limb measurements. These exist for certain time periods but as far as we know are not included in operational assimilation products.

[11]Data assimilation which is continuous in time (4D-var) most likely reduces the problem of aliasing due to asynoptic sampling (see footnote 2). This, however, needs to be tested.

correlations in the final retrieval come either from this matrix, or from the constraint field itself, which is a forward integration in time of earlier wave observations (i.e. vertical wave propagation dynamics). Our results suggest that it is best to have a diagonal error covariance matrix, and let all the vertical correlations in the constraint come from the model dynamics. Even if the error covariance matrix is diagonal, the assimilation does not necessarily resolve the waves during a final or sudden warming. The main question is does the model simulate the dynamics correctly, and if not, can the radiances constrain it to do so? Our results indicate that TOVS observations are not able to constrain the model to sharp temperature features.

It is important to keep in mind that since our analysis is very idealized, our results are a lower bound on retrieval errors. Many errors which affect real observations are absent from our setup. Some examples include errors due to clouds, to the interpolation from the satellite tracks to the operational grid, and to uncertainties in the weighting functions (see Smith et al. 1979; Kidwell, 1986; Kidder, 1995). Our results can be regarded as a determination of the information content of TOVS weighting functions, and the implications for resolving wave structure and evolution. The relevance to assimilation products is in indicating what features of a forecast model the TOVS measurements are able to constrain, and what features not.

Finally, even though the quantitative results of this study will not apply for future observing systems, the qualitative results should still be relevant. Also, our approach, of simulating a simplified retrieval system in a model, is simple enough to be used as a test of the limitations of these new observations for studying the dynamical phenomena of interest.

Acknowledgments: We thank Thomas Kleespies from NESDIS for supplying the OPTRAN code. We thank Larry McMillin, Philip Rosenkranz, and Laurie Rokke

for answering questions regarding satellite retrievals. This research was supported by grants ATM9813795 from the National Science Foundation, NAG5-5147 from National Aeronautics and Space Administration, and DEFG02-93ER61673 from the Department of Energy.

Appendix

A The model generated temperature fields

We use both steady state and time dependent quasi-geostrophic β-plane channel models, linearized around a zonal mean basic state that has a stratospheric jet characteristic of winter, and a US standard winter stratosphere temperature profile. The model is forced with a wave 1 or 2 (we only show results for wave 1). The zonal mean is calculated from the basic state temperature $T(y_o, z)$, assuming a thermal wind balance:

$$T(y,z) = -\int_{y_o}^{y} \frac{f_o T_o}{g} U_z dy + T(y_o, z) \tag{2}$$

where z is log pressure, y is latitude, f_o is the Coriolis parameter, g the gravitational acceleration, and T_o a reference temperature.

The temperature wave field is calculated from the model calculated geopotential height perturbation (φ): $T' = \frac{T_o}{g}\frac{\partial \varphi}{\partial z} = g\frac{T}{T_o}$. We characterize the wave by its amplitude and phase (assuming wave 1): $T'(y,z) = A \cdot a(y,z)\cos(x + \alpha(y,z))$ where a and α are the amplitude and phase of the wave, respectively, x is longitude, and A is an arbitrary amplitude factor (the model is linear), which we specify. In all figures latitude is nondimensionalized by radii deformation ($1272.5km$), and since most of our model studies are of the southern hemisphere, latitude starts at the pole and the equator is at large y.

B The calculation of TOVS radiances

We use the OPTRAN fast transmittance model (McMillin et al. 1995) to calculate the weighting functions of the HIRS and MSU channels, and equations 3-4 of Taylor et al. (1972) for the SSU channels. A detailed description of TOVS instruments can be found in Smith et al. (1979), Kidder and Vonder Haar (1995) and Kidwell (1986).

Since we are interested in stratospheric waves, the lower boundary of our model is at 14km (roughly tropopause height), where we assume:

$$I(z_s) = B(T(z_s))\epsilon_\nu \tag{3}$$

Such a relation is generally used for the ground, where ϵ_ν is an effective emissivity, and a gray or black body law is used. The justification for using equation 3 at 14km is that it has little effect on the retrievals in the stratosphere, while keeping the equations simple. Out of the 7 channels we use, the bottom boundary condition affects only the two lowest channels, since the transmittance of the other 5 channels is less than 0.1 at the bottom. We test the sensitivity of the retrievals to ϵ_ν by varying it between 0.5-1.5, and find it to be quite small, with differences being only at the surface, of less than 2 K[12]. For simplicity, we use $\epsilon_\nu = 1$.

For some of the runs we add noise to the radiances. We use white noise with a mean of zero, and a standard deviation that is characteristic of actual instrument errors. Table 1 lists the instrumental errors of the channels used in this study, both in terms of the temperature error (taken from the NOAA POD guide, Kidwell, 1986) and in terms of the corresponding radiance error, for the channel mean scene temperature

[12]The differences are 2 K for the more sensitive Chahine retrieval, and are much smaller for the Minimum Variance retrieval which is closer to the operational retrievals (see section 2.2).

(also taken from the NOAA POD guide) and a wavenumber of 668 cm^{-1} for all channels. We do not use instrument radiance errors directly because we do not use real channel frequencies (see section 2.1) and our radiances are different from the real ones.

C The Minimum Variance method

Our derivation follows Rodgers (1976), with slightly different notation. Using a single frequency for all channels, we solve for the Planck function. Equation 1 in a matrix form is:

$$\mathbf{r} = \mathbf{K}\mathbf{b} \tag{4}$$

$\mathbf{r}$ and $\mathbf{b}$ represent the set ot radiance measurements by the satellite channels and the discretized vertical profile of the Planck function, respectively, and $\mathbf{K}$ is the forward operator relating the Planck function to radiances.

The MV retrieval consists of solving the following equation:

$$\hat{\mathbf{b}} = \mathbf{b_o} + \mathbf{S_o}\mathbf{K}^T(\mathbf{K}\mathbf{S_o}\mathbf{K}^T + \mathbf{S}_\epsilon)^{-1}(\mathbf{r} - \mathbf{K}\mathbf{b_o}) \equiv \mathbf{b_o} + \mathbf{A_{km}}(\mathbf{b} - \mathbf{b_o}) \tag{5}$$

where $(\ )^T$ denotes a transpose. $\hat{\mathbf{b}}$ is the retrieved profile, $\mathbf{b}$ is the *inverse solution* (equation 4), and $\mathbf{b_o}$ the *constraint profile*. $\mathbf{S_o}$ is the constraint error covariance matrix, and $\mathbf{S}_\epsilon$ is the radiance error covariance matrix which is diagonal (meaning the instrument errors are uncorrelated between the different channels)[13]. The error covariance matrix of the retrieval solution $\hat{\mathbf{b}}$ is $\hat{\mathbf{S}} = (\mathbf{S_o}^{-1} + \mathbf{K}^T\mathbf{S}_\epsilon{}^{-1}\mathbf{K})^{-1}$.

$\mathbf{A_{km}}$ is called the *Averaging Kernel Matrix*, because the retrieval at a given height is an average of the whole profile weighted by the corresponding row of $\mathbf{A_{km}}$ (Backus and Gilbert, 1970; Conrath, 1972). The columns of $\mathbf{A_{km}}$ are the set of *response functions*, which are essentially the response of the observing system to a spike of temperature perturbation introduced at the different levels (figure 5).

[13]Note that the error covariance matrix of the inverse solution $\mathbf{b}$ equals $\mathbf{S_r} = (\mathbf{K}^T\mathbf{S}_\epsilon{}^{-1}\mathbf{K})^{-1}$, which is not diagonal.

The constraints are constructed in the following manner:

The diagonal constraint: We use a diagonal $\mathbf{S_o}$, and in most cases, $\mathbf{b_o}$ is specified to be the zonal mean basic state at the middle latitude.

The non-diagonal constraint: Given a climatology of L profiles $\mathbf{x_l}$, where the subscript l denotes the measurement number, and $\overline{\mathbf{x}}$ is the mean of this sample, we choose $\mathbf{b_o} = \overline{\mathbf{x}}$ and $\mathbf{S_o} = \frac{1}{L}\sum_{l=1}^{L}\left[(\mathbf{x_l} - \overline{\mathbf{x}})(\mathbf{x_l} - \overline{\mathbf{x}})^T\right]$. We use 50 days of integration of our time dependent model where we superpose a zonaly propagating wave 1 (phase speed of $15\frac{m}{sec}$) and a stationary wave 1, to construct the climatology from which we calculate the constraint. The wave field changes periodically (with a period of 20.6 days), due to the different phase superpositions of the standing and traveling waves. The control experiments use an amplitude factor of $A = 4$ (see appendix A) and all the grid points between $y = 1 - 4$ $(1272.5 - 5090km)$ for the climatology. Variations on the control run include choosing a different latitude range from which to construct the climatology and choosing a different wave amplitude factor (ranging from 0.0 to 8.0).

D References

Andersson, E., J. Pailleux, J.-N. Thepaut, J. R. Eyre, A. P. McNally, G. A. Kelly, and P. Courtier, 1994. Use of cloud-cleared radiances in three/four-dimensional variational data assimilation. *Quart. J. Roy. Meteor. Soc.* **120**: 627-653.

Andrews, D. G., J. R. Holton and C. B. Leovy, 1987. *Middle atmosphere dynamics.* Academic Press, 489pp.

Backus, G. E., and J. F. Gilbert, 1970. Uniqueness in the inversion of inaccurate gross earth data. *Philos. Trans. R. Soc. London, Ser A,* **266**: 123-192.

Chahine, M. T., 1968. Determining of the temperature profile in an atmosphere from its outgoing radiance. *J. Opt. Soc. Amer.* **58**: 1634-1637.

Chahine, M. T., 1970. Inverse problems in radiative transfer: A determination of atmospheric parameters. *J. Atmos. Sci.* **27**: 960-967.

Charney, J. G., and P. G. Drazin, 1961. Propagation of planetary scale disturbances from the lower into the upper atmosphere. *J. Geophys. Res.* **66**: 83-110.

Conrath, B. J., 1972. Vertical resolution of temperature profiles obtained from remote radiation measurements. *J. Atmos. Sci.* **29**: 1262-1271.

Derber, J. C., and W.-S. Wu, 1998. The use of TOVS cloud-cleared radiances in the NCEP SSI analysis system. *Mon. Wea. Rev.* **126**: 2287-2299.

Dickinson R. E., 1968. Planetary waves propagating vertically through weak westerly wind wave guides. *J. Atmos. Sci.* **25**: 984-1002.

Dickinson R. E., 1969. Vertical propagation of planetary Rossby waves through and atmosphere with Newtonian cooling. *J. Geophys. Res.* **74**: 929-938.

Eyre, J. R., G. A. Kelly, A. P. McNally, E. Andersson, and A. Persson, 1993. Assimilation of TOVS radiance information through one-dimensional variational analysis. *Quart. J. Roy. Meteor. Soc.* **119**: 1427-1463.

Graves, D. S., 1986. Evaluation of satellite sampling of the middle atmosphere using the GFDL SKYHI general circulation model. *Ph.D. dissertation.* Princeton university. 314 pp.

Harnik, N., and R. S. Lindzen, 2001: The effect of reflecting surfaces on the vertical structure and variability of stratospheric planetary waves. *J. Atmos. Sci.* , in press.

Hirota, I., and Y. Sato, 1969. Periodic variation of the winter circulation and intermittent vertical propagation of planetary waves. *J. Meteorol. Soc. Jpn.* **47**: 390-402.

Karoly, J. D., and Graves, D. S., 1990. On data sources and quality for the southern hemisphere stratosphere. *Dynamics, transport and photochemistry in the middle atmosphere of the southern hemisphere.* (A. O'Neill ed.): 19-32. Kluwer Academic Publishers.

Kidder, S. Q. and T. H. Vonder Haar, 1995. *Satellite meteorology, an introduction.* Academic Press. 466pp.

Kidwell, K. B., 1986. *The NOAA Polar Orbiter Data user's guide.* NOAA/NESDIS satellite data services division, Washington DC;
An updated version: http://www2.ncdc.noaa.gov:80/docs/podug/.

Liou, K.-N., 1980. *An introduction to atmospheric radiation.* Academic Press. 392pp.

Mateer, C. L., 1965. On the information content of Umkehr observations. *J. Atmos. Sci.* **22**: 370-381.

McMillin, L. M., L.J. Crone, M. D. Goldberg, and T. J. Kleespies, 1995. Atmospheric transmittance of an absorbing gas. 4. OPTRAN: a computationally fast and accurate transmittance model for absorbing gases with fixed and with variable mixing ratios at variable viewing angles. *Appl. Opt.* **34**: 6269-6274.

Miles, T. and A. O'Neill, 1989. *Comparison of satellite derived dynamical quantities in the stratosphere of the southern hemisphere.* Proceedings of MASH workshop, Williamsburg, VA.

Rodgers, C. D., 1976. Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation. *Rev. Geophys. and Space Phys.* **14**: 609-624.

Rodgers, C. D., 1990. Characterization and error analysis of profiles retrieved from remote sounding measurements. *J. Geophys. Res.* **95**: 5587-5595.

Salby, M. L., 1982. Sampling theory for asynoptic satellite observations. Part I: Space-time spectra, resolution and aliasing. *J. Atmos. Sci.* **39**: 2577-2600.

Smith, W. L., and H. M. Woolf, 1976. The use of statistical covariance matrices for interpreting satellite sounding radiometer observations. *J. Atmos. Sci.* **33**: 1127-1140.

Smith, W. L., H. M. Woolf, and C. M. Hayden, 1979. The TIROS-N operational vertical sounder. *Bull. Am. Meteorol. Soc.* **58**: 1177-1187.

Stephens, G. L., 1994. *Remote sensing of the lower atmosphere : an introduction.* Oxford University Press. 523pp.

Taylor, F. W., J. T. Houghton, G. D. Peskett, C. D. Rodgers, and E. J. Williamson, 1972. Radiometer for remote sounding of the upper atmosphere. *Appl. Opt.* **11**: 135-141.

Table Captions:

Table 1. The channels used in this study. Error data is taken from the NOAA POD guide (Kidwell, 1986), see appendix B for details. A scale height of 7km is used to calculate the height of the peaks of the SSU channels from the corresponding pressures, and the pressures of the peaks of the HIRS channels from the corresponding heights.

Table 2. The variance of the constraint of the nondiagonal MV retrieval (see text for details).

Figure Captions:

Figure 1. The weighting functions that peak in our model domain. Solid lines are the channels used in our study (marked next to the peaks) and the dashed line is the MSU channel we drop. In all relevant figures, the grid lines below 49 km are at the peaks of the weighting functions.

Figure 2. Left: A 'true' temperature profile (line with solid circles), and various retrievals of it: the Chahine retrieval (circles), and diagonal MV retrievals using the constraint profile shown (thin dotted line) and a constant variance of 5 K (dash-dot), 10 K (solid) and 20 K (dashed). Right: The true minus retrieved and true minus constraint profiles shown on the left (using same line types).

Figure 3. The 10 largest eigenvalues of the Averaging Kernel Matrix for a diagonal MV retrieval, with a constant variance of 2.5 K (asterisks), 5 K (circles), 10 K (stars),20 K (diamonds) and 40 K (plusses).

Figure 4. The first six eigenvectors of the Averaging Kernel Matrix for a diagonal MV retrieval using a constant variance of 5 K (dash-dot), 10 K (solid) and 20 K (dashed).

Figure 5. The response functions to a spike perturbation of temperature at various heights (marked at the top of each sub-figure), for a non-diagonal (solid) MV retrieval and the corresponding diagonal (dashed) retrieval (see text for details).

Figure 6. The 'simple' wave run: **a.** A latitude-height plot of the model generated stationary wave 1 temperature amplitude (solid) and phase (dashed, units of π), for the case of a wave that propagates up throughout the stratosphere. **b.** The damping rate used in the simple run of a (solid) and in the damped run

of figure 7b (dashed) (day^{-1}). **c.** A diagonal MV retrieval of a with a constant variance of 7°K. **d.** The difference between the true and diagonally retrived temperature amplitudes (**a**-**c**). **e**. and **f**. As in **c** and **d** respectively, but for the nondiagonal retrieval of figure 5.

Figure 7. The undamped vs damped wave run: Longitude-height plots of temperature (°K) at $y = 2.3$ of: **a.** The model generated wave field of figure 6a (using the damping profile in figure 6b, solid line). **b.** Same as **a**, only for a run with stratospheric damping (figure 6b, dashed line). **c.** The undamped minus damped model waves (**a**-**b**). **d.** and **e.** The diagonal retrievals of **a** and **b**, respectively, using a constant variance of 7°K. **f.** The difference of the diagonally retrieved fields (**d**-**e**). **g.** and **h.** The nondiagonal retrieval of **a** and **b**. **i.** As in **f**, but for the nondiagonal retrieval. Dashed lines are zero.

Figure 8. Latitude height sections of zonal mean wind (top) and wave 1 temperature amplitude (bottom, solid) and phase (bottom, dashed) for a model run with partial reflection due to turning surfaces (left) and a run with a critical surface (right). The shading (top left) shows the evanescent region, between the two turning surfaces. The thick line (top right) shows the critical surface.

Figure 9. As in figure 7, only for the partially reflected wave (denoted by T_{TS}, left row), on the basic state of figure 8(left), and an absorbed wave (denoted by T_{CS}, middle row), on the basic state of figure 8(right). Difference fields are for the partially reflected minus the absorbed wave.

Figure 10. The 'true' temperature field generated by the model with a summer basic state (top) and its non-diagonal MV retrieval (bottom). Left: Wave number 1 temperature amplitude. Right: A longitude-height section of the temperature

wave number 1, at a latitude of y=2.83. The non-diagonal constraint is the same as in figure 5.

Figure 11. Longitude-height sections, at $y = 2.3$ for three days (6, 7, and 8, from right to left) of the model generated (top) wave field and its nondiagonal retrieval (bottom), for the time dependent run described in the text. Contour interval is 4°K, and zero lines are dashed.

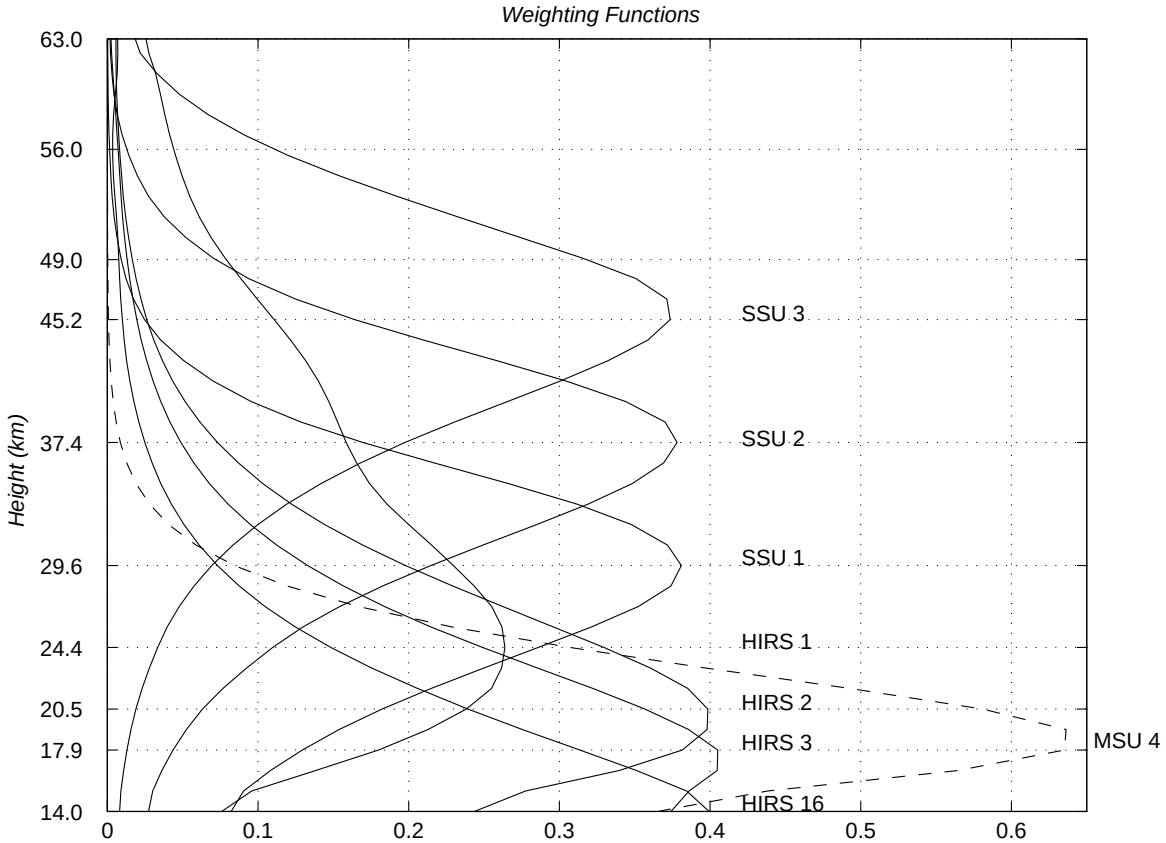

Figure 1: The weighting functions that peak in our model domain. Solid lines are the channels used in our study (marked next to the peaks) and the dashed line is the MSU channel we drop. In all relevant figures, the grid lines below 49 km are at the peaks of the weighting functions.

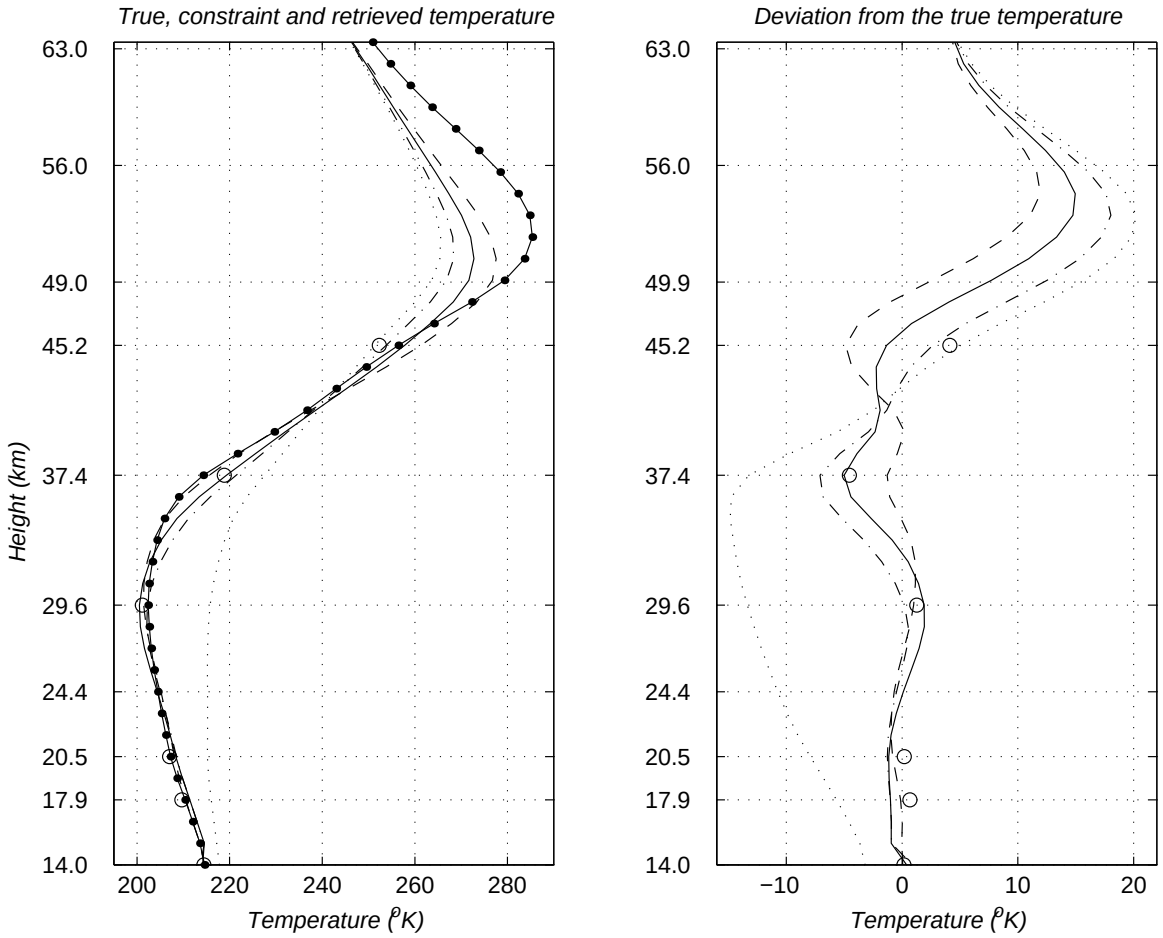

Figure 2: Left: A 'true' temperature profile (line with solid circles), and various retrievals of it: the Chahine retrieval (circles), and diagonal MV retrievals using the constraint profile shown (thin dotted line) and a constant variance of 5 K (dash-dot), 10 K (solid) and 20 K (dashed). Right: The true minus retrieved and true minus constraint profiles shown on the left (using same line types).

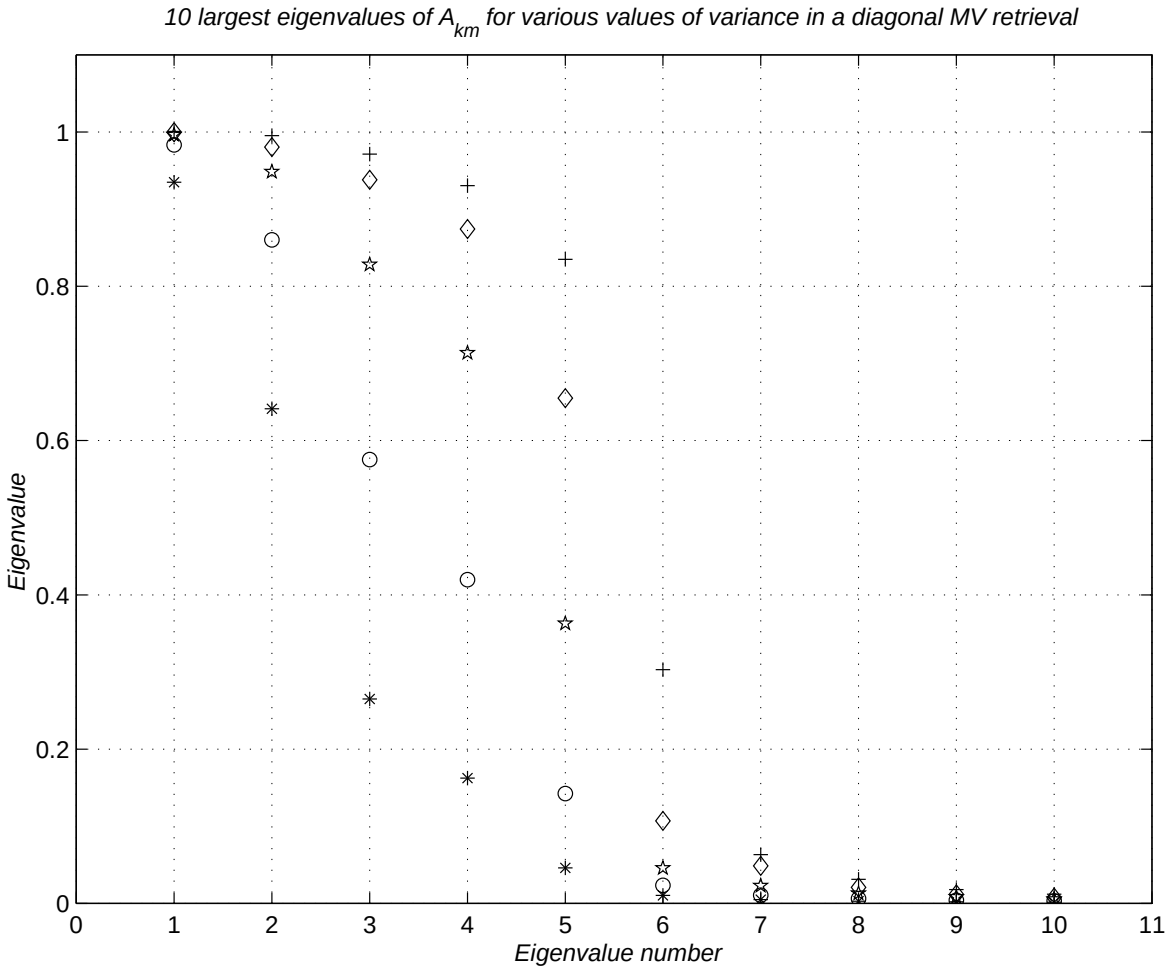

Figure 3: The 10 largest eigenvalues of the Averaging Kernel Matrix for a diagonal MV retrieval, with a constant variance of 2.5 K (asterisks), 5 K (circles), 10 K (stars),20 K (diamonds) and 40 K (plusses).

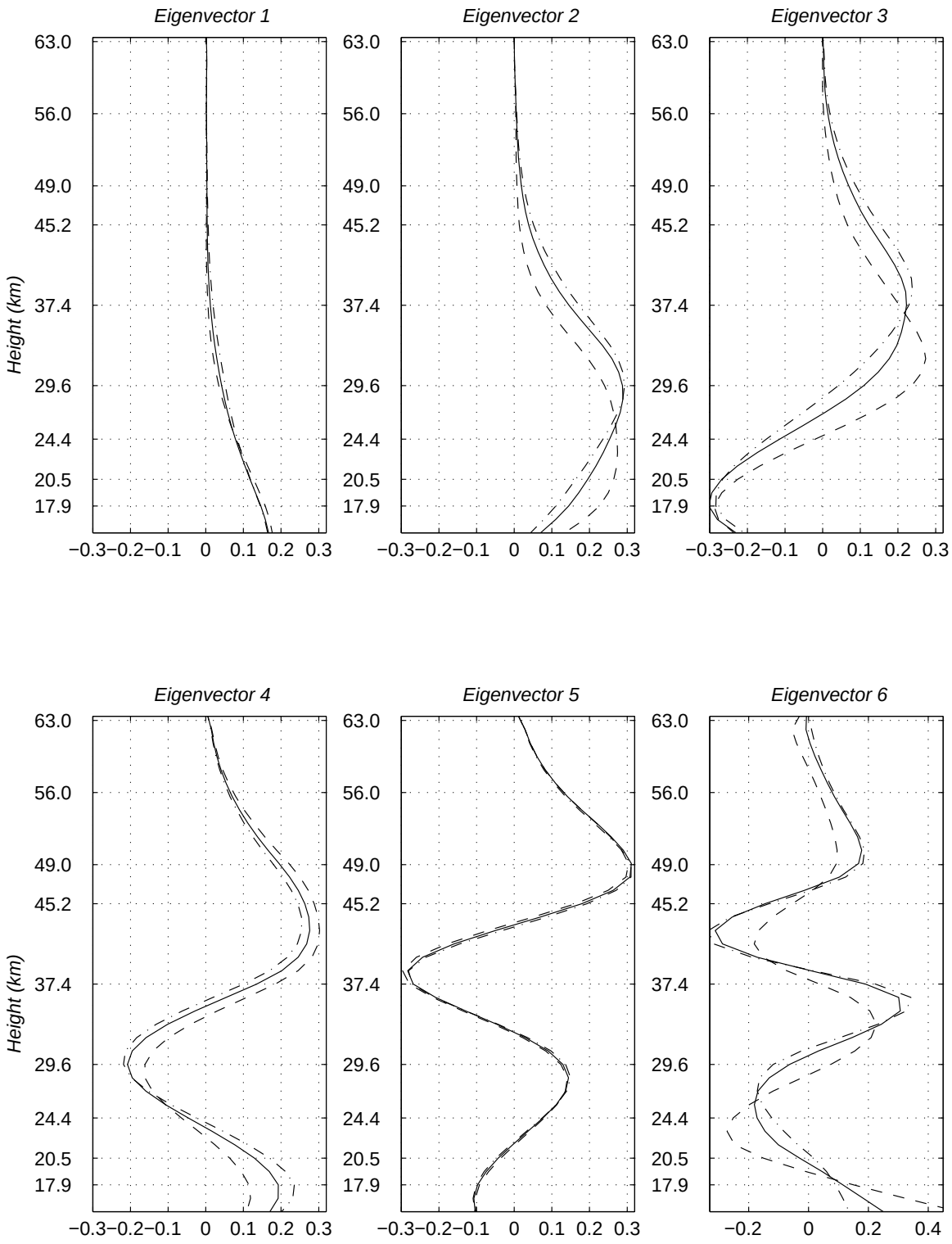

Figure 4: The first six eigenvectors of the Averaging Kernel Matrix for a diagonal MV retrieval using a constant variance of 5 K (dash-dot), 10 K (solid) and 20 K (dashed).

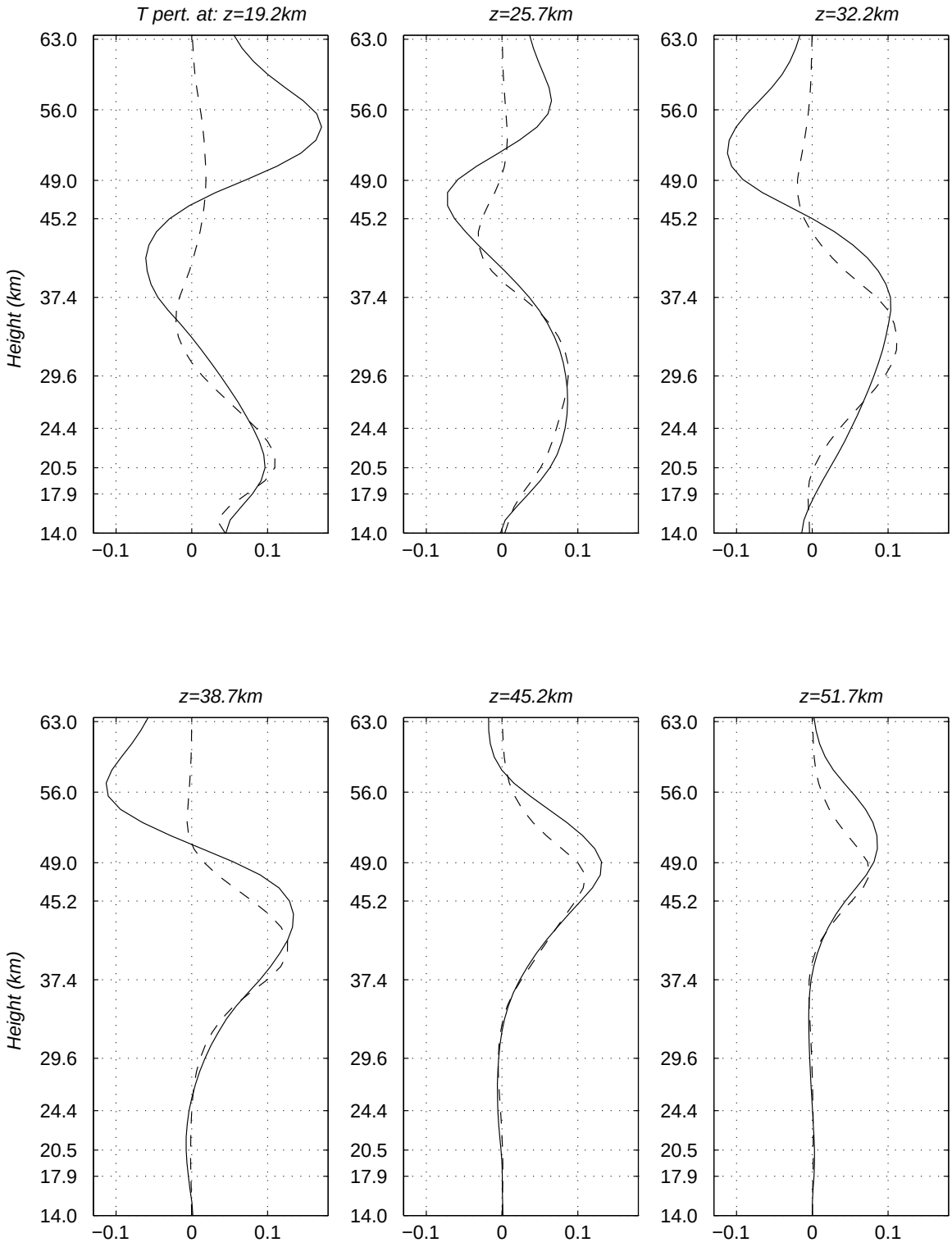

Figure 5: The response functions to a spike perturbation of temperature at various heights (marked at the top of each sub-figure), for a non-diagonal (solid) MV retrieval and the corresponding diagonal (dashed) retrieval (see text for details).

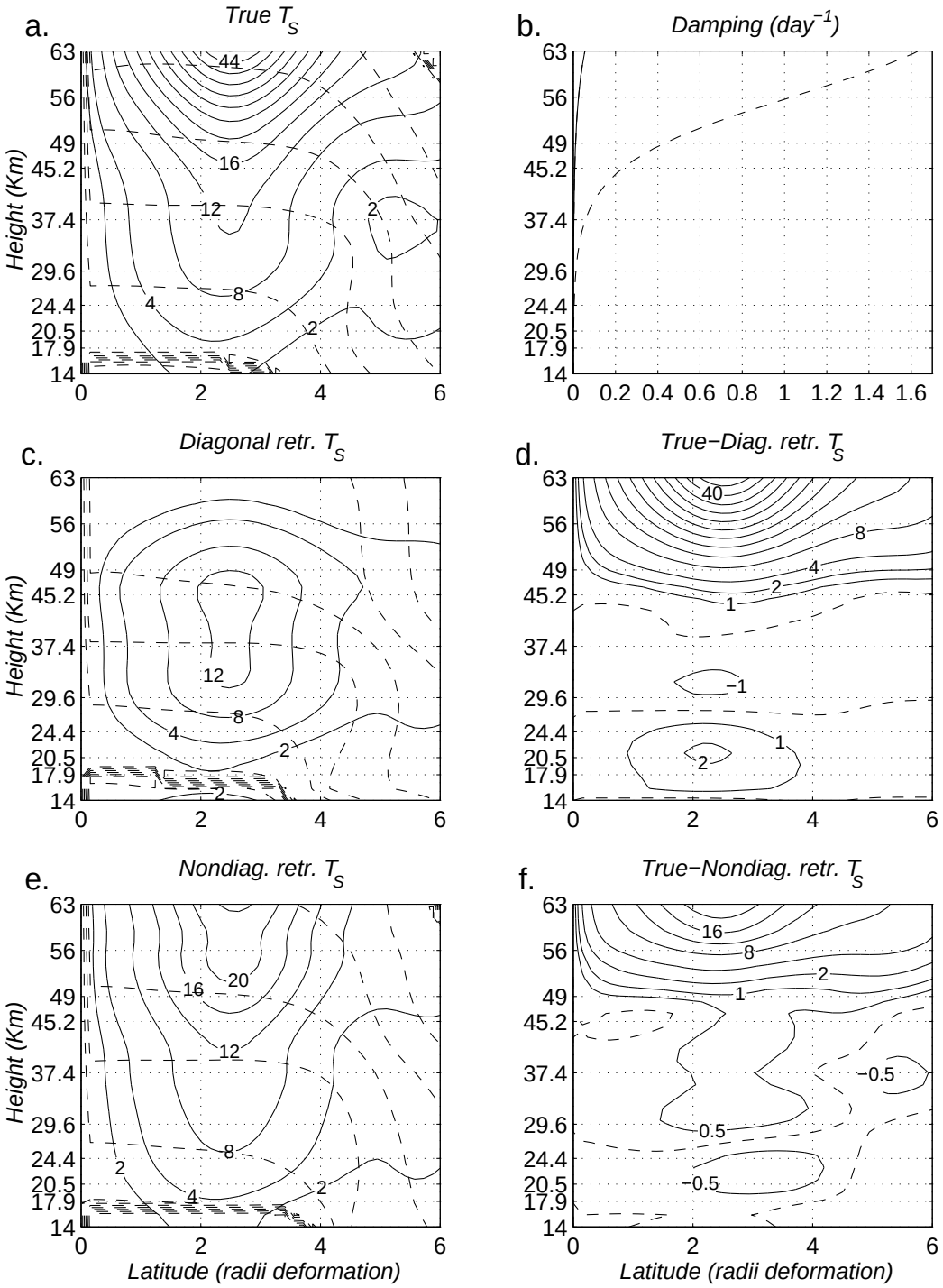

Figure 6: The 'simple' wave run: **a.** A latitude-height plot of the model generated stationary wave 1 temperature amplitude (solid, °K) and phase (dashed, intervals of 1 radian), for the case of a wave that propagates up throughout the stratosphere. **b.** The damping rate used in the simple run of a (solid) and in the damped run of figure 7b (dashed) (day^{-1}). **c.** A diagonal MV retrieval of a with a constant variance of 7°K. **d.** The difference between the true and diagonally retrived temperature amplitudes (**a**-**c**). **e**. and **f**. As in **c** and **d** respectively, but for the nondiagonal retrieval of figure 5.

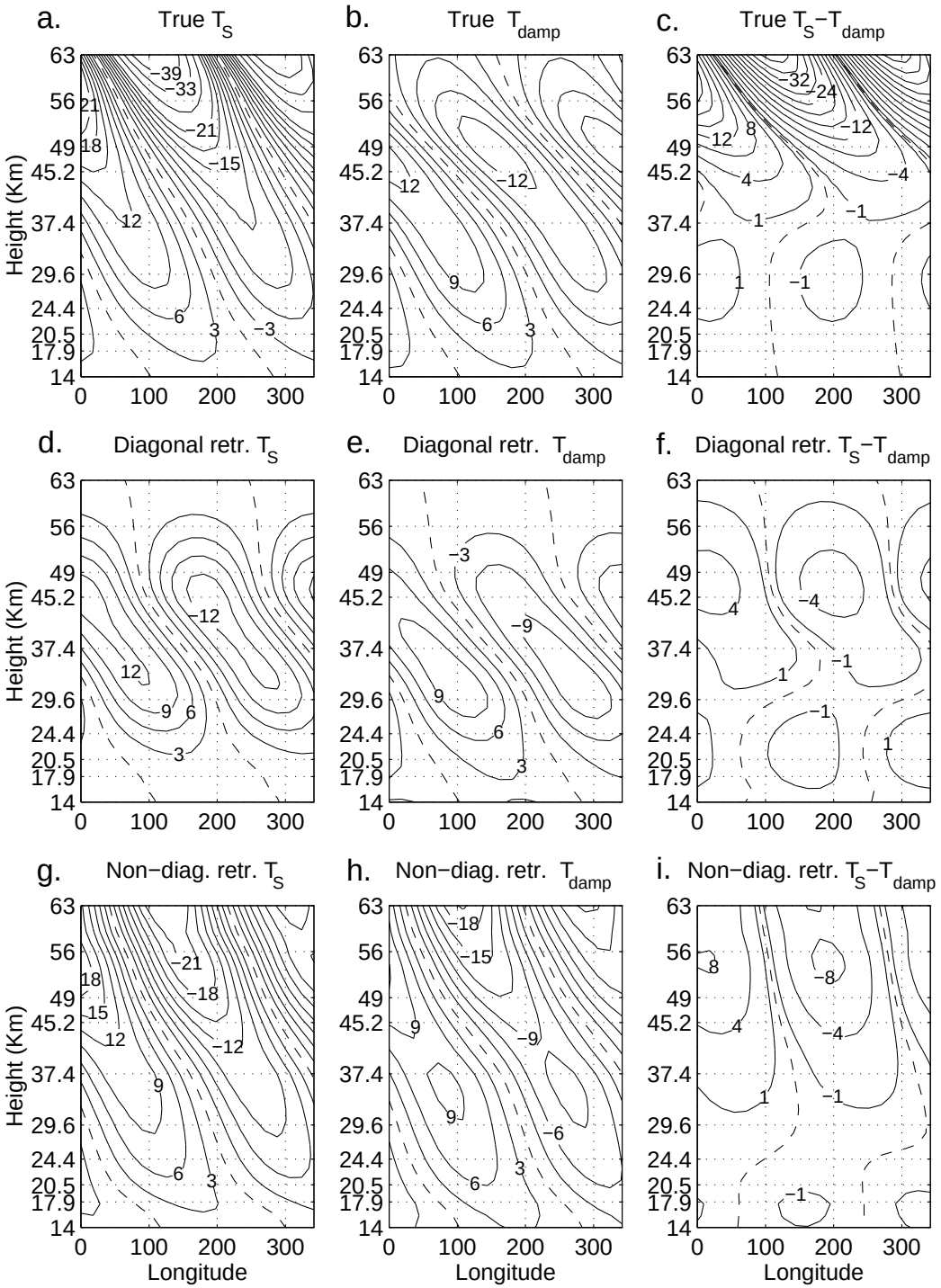

Figure 7: The undamped vs damped wave run: Longitude-height plots of temperature (°K) at $y = 2.3$ of: **a.** The model generated wave field of figure 6a (using the damping profile in figure 6b, solid line). **b.** Same as **a**, only for a run with stratospheric damping (figure 6b, dashed line). **c.** The undamped minus damped model waves (**a**-**b**). **d.** and **e.** The diagonal retrievals of **a** and **b**, respectively, using a constant variance of 7°K. **f.** The difference of the diagonally retrieved fields (**d**-**e**). **g.** and **h.** The nondiagonal retrieval of **a** and **b**. **i.** As in **f**, but for the nondiagonal retrieval. Dashed lines are zero.

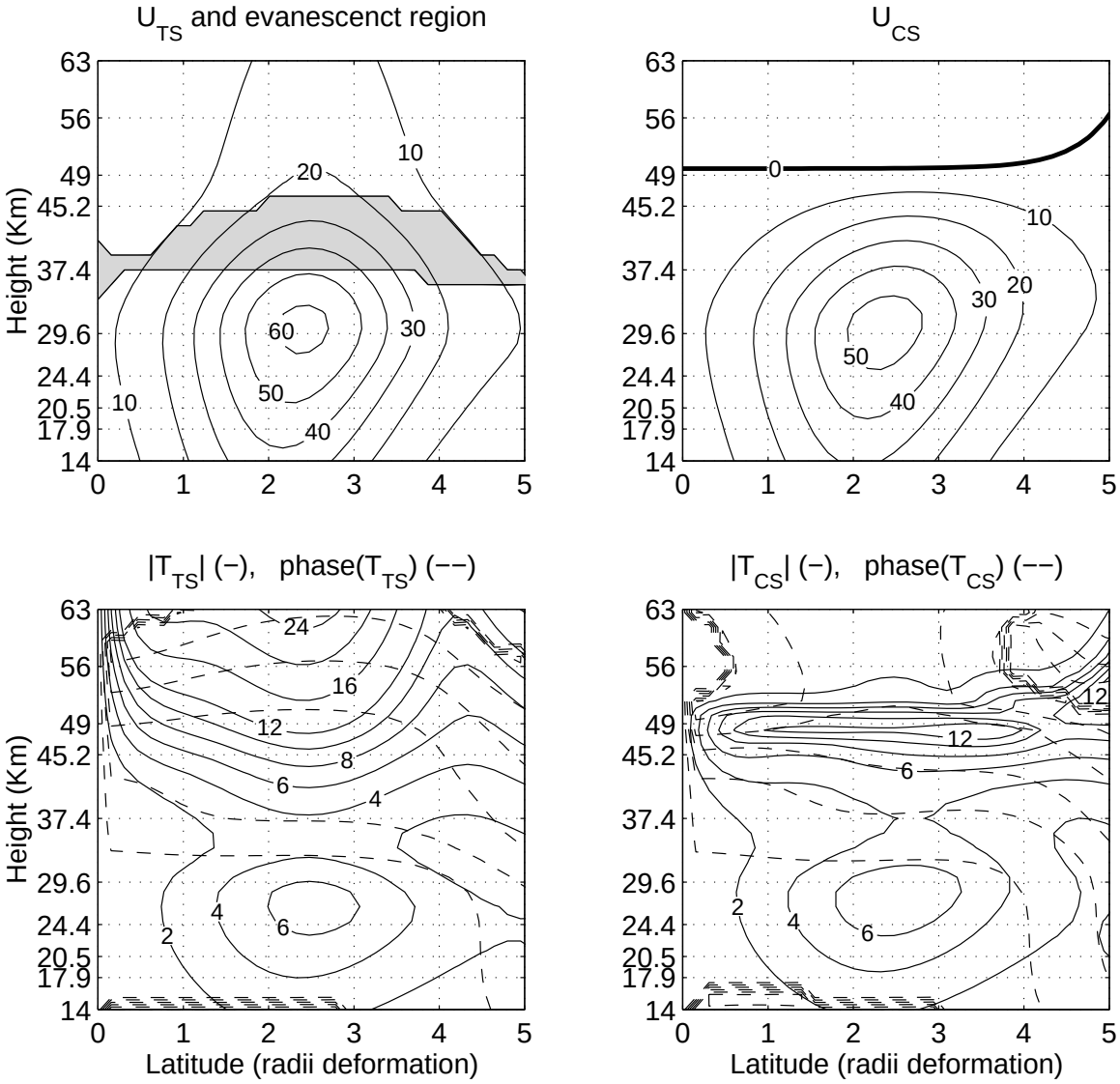

Figure 8: Latitude height sections of zonal mean wind (top, $msec^{-1}$, contour interval of 10) and wave 1 temperature amplitude (bottom, solid, °K, contour interval 2 below 12, and 4 above) and phase (bottom, dashed, contour interval of 1 radian) for a model run with partial reflection due to turning surfaces (left) and a run with a critical surface (right). The shading (top left) shows the evanescent region, between the two turning surfaces. The thick line (top right) shows the critical surface.

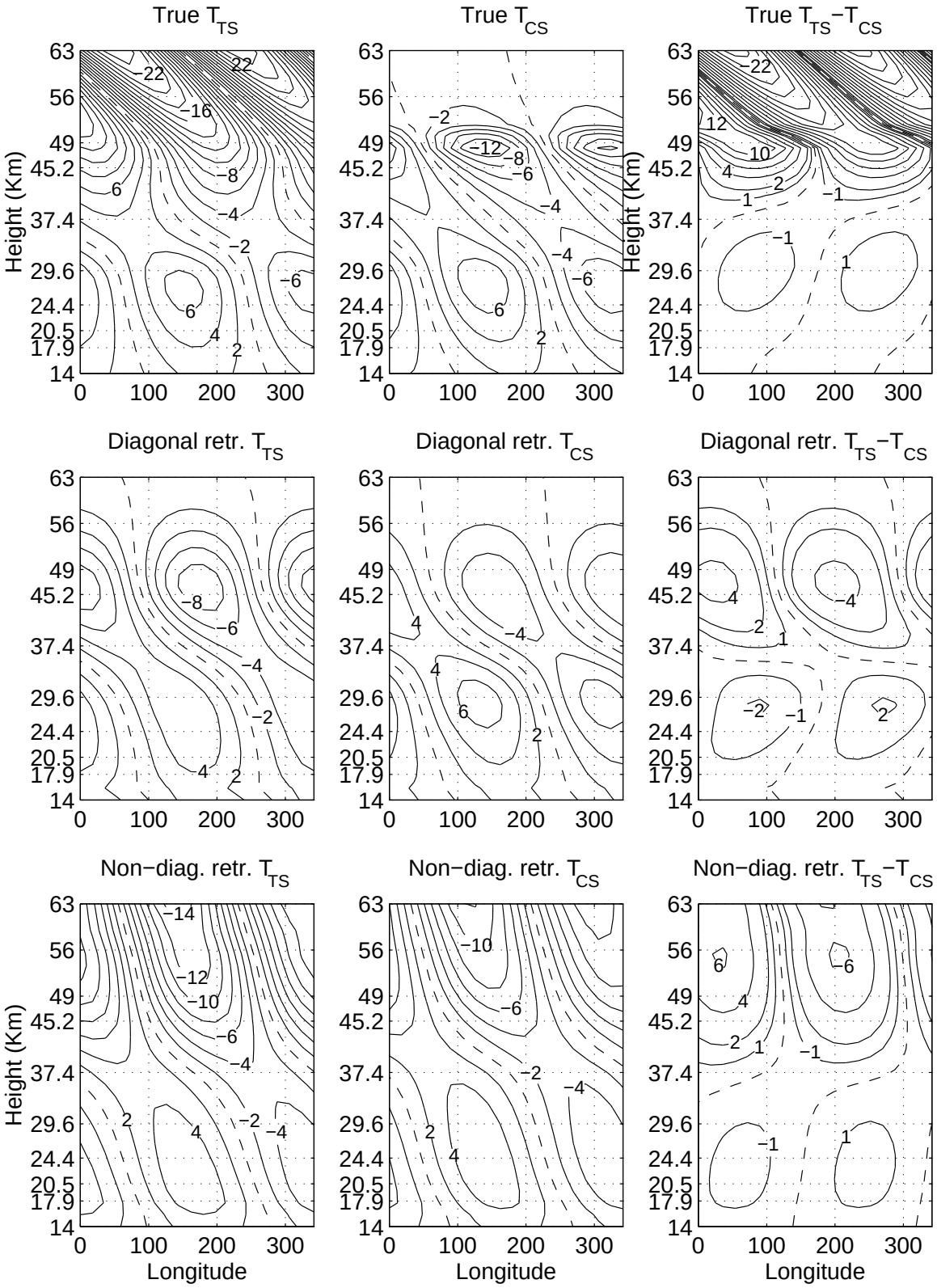

Figure 9: As in figure 7, only for the partially reflected wave (denoted by T_{TS}, left row), on the basic state of figure 8(left), and an absorbed wave (denoted by T_{CS}, middle row), on the basic state of figure 8(right). Difference fields are for the partially reflected minus the absorbed wave.

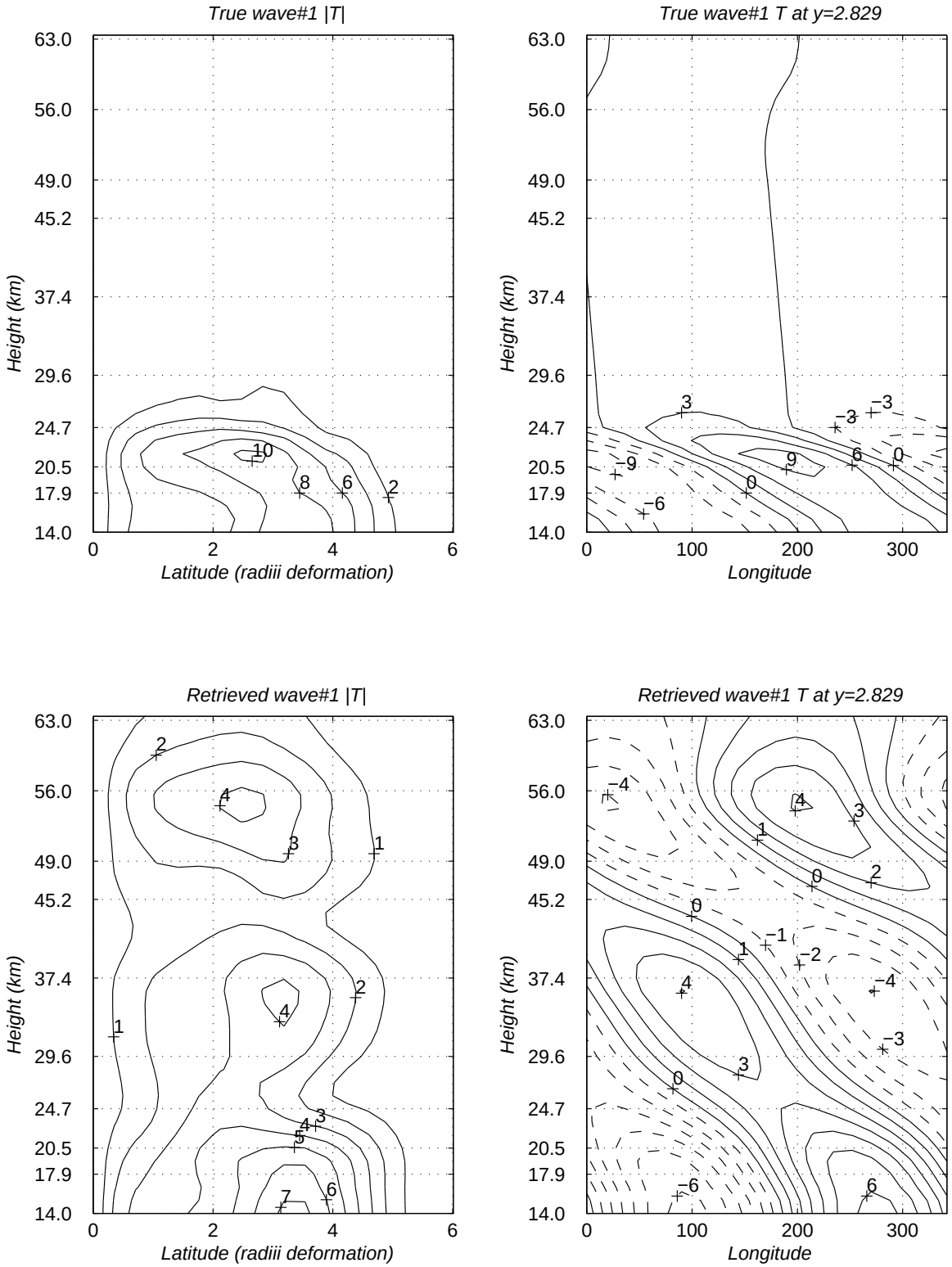

Figure 10: The 'true' temperature field generated by the model with a summer basic state (top) and its non-diagonal MV retrieval (bottom). Left: Wave number 1 temperature amplitude. Right: A longitude-height section of the temperature wave number 1, at a latitude of y=2.83. The non-diagonal constraint is the same as in figure 5.

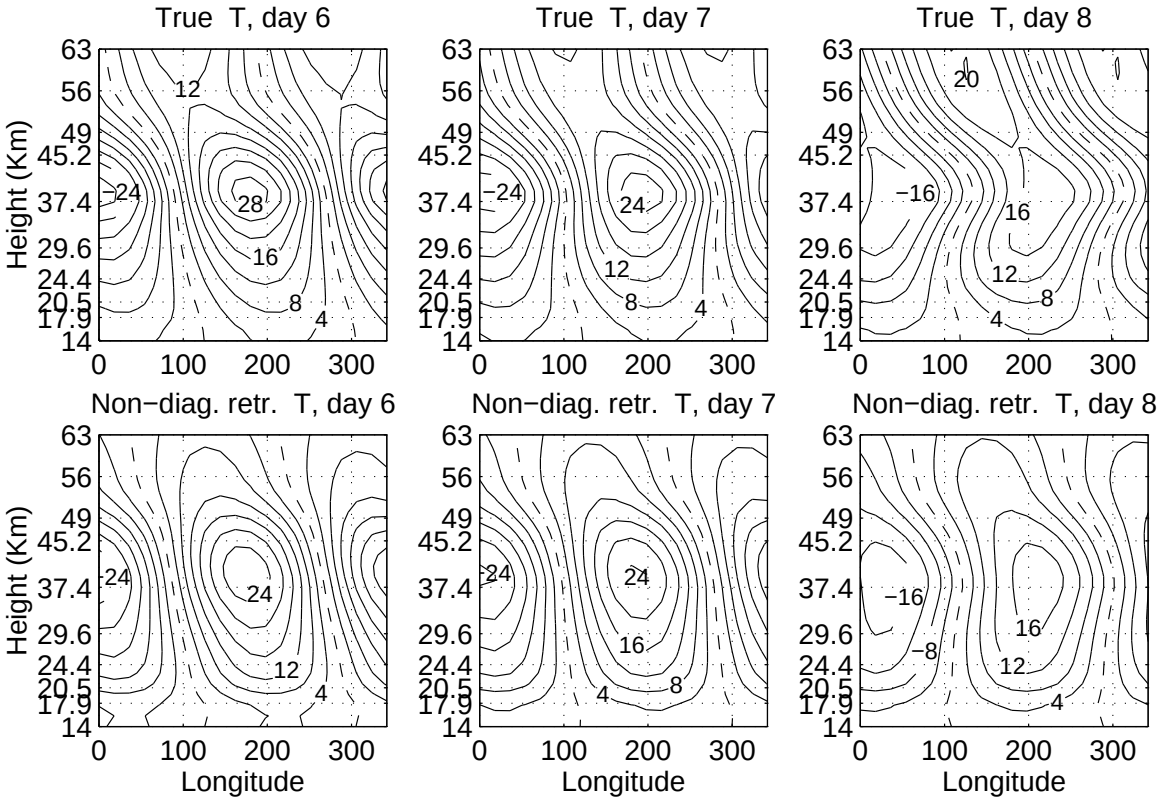

Figure 11: Longitude-height sections, at $y = 2.3$ for three days (6, 7, and 8, from right to left) of the model generated (top) wave field and its nondiagonal retrieval (bottom), for the time dependent run described in the text. Contour interval is 4°K, and zero lines are dashed.

Instrument and channel	Weighting function peak km (hPa)	ΔT(K) at mean scene temperature	Mean scene temperature (K)	ΔI $10^{-5}\frac{W/m^2}{Sr/m}$	Instrument wavenumber (cm^{-1})
HIRS 16	14.0(135.3)	0.31	230.0	0.32	2265.5
HIRS 3	17.9(77.5)	0.55	220.0	0.51	689.22
HIRS 2	20.5(53.5)	0.74	220.0	0.68	682.22
HIRS 1	24.4(30.6)	2.77	235.0	2.96	668.51
SSU 1	29.4(15)	0.25	273.0	0.36	668.00
SSU 2	37.1(5)	0.50	273.0	0.72	668.00
SSU 3	45.5(1.5)	1.25	273.0	1.79	668.00

Table 1: The channels used in this study. Error data is taken from the NOAA POD guide (Kidwell, 1986), see section B for details. A scale height of 7km is used to calculate the height of the peaks of the SSU channels from the corresponding pressures, and the pressures of the peaks of the HIRS channels from the corresponding heights.

Height (km)	Variance (K)
14.0	3.2
20.5	10.2
27.0	11.2
33.5	11.7
40.0	12.5
46.5	13.4
53.0	11.0
59.5	5.1

Table 2: The variance of the constraint of the nondiagonal MV retrieval (see text for details).