5. Zonal-mean detection studies: Sensitivity to observational, model and pre-formatting uncertainties

There have been a number of previous climate change detection studies that have considered observed changes in zonally-averaged upper-air temperatures. The earliest studies used pattern correlation techniques (Santer et al., 1996a, Tett et al., 1996, for example). More recent studies have used optimal regression techniques, either with a fixed signal (AT99, Tett et al., 2001), or a more complex spatio-temporal input diagnostic (Hill et al., 2001, G. S. Jones et al., 2001). All such studies have consistently found a detectable anthropogenic influence and, more tentatively, in some cases, where natural (external) signals are considered, solar and volcanic forcing influences. In none of these studies has the sensitivity of the results to the details of the precise methodology employed been considered in a rigorous manner. In this chapter, the sensitivity of fixed-signal zonally-averaged optimal regression detection results to the likely major sources of uncertainties is considered. The sources of uncertainty explicitly considered here are:

- Pre-processing choices of modelled, and observed, temperature fields.
- Uncertainty in the veracity of the observational upper air temperature dataset.
- Uncertainty in the GCM model fields.
- Whether the recently-observed warming in the troposphere, **and** cooling in the stratosphere, are both included in the detection exercise.
- Explicitly accounting for the presence of additional uncertainty in the model signals due to finite ensemble size.

In section 5.1 methods used to address these potential sources of uncertainty are outlined. Section 5.2 presents a summary of results, detailing the sensitivity to the criteria highlighted in section 5.1, whilst section 5.3 concludes.

5.1 Identifying and testing for sources of uncertainty in fixed-signal zonally-averaged optimal regression detection studies

Input diagnostics in the current study are calculated in the same manner as those of previous fixed-signal zonal-mean optimal regression detection studies (AT99, Tett et al., 2001). Annual-mean temperature values are zonally averaged at eight standard WMO standard reporting pressure levels (850, 700, 500, 300, 200, 150, 100, and 50 hPa) for both modelled and observed fields, with model data masked to match the spatio-temporally varying observed data availability. Subsequently, a difference field is constructed based on the 1985-1995 average minus the 1960-1980 average temperatures. The sole criterion for inclusion, is the existence of a single annual value in each period, for any given grid-box. Using such liberal criteria to derive the individual period means will have obvious implications in terms of adding noise to the difference field in poorly sampled regions. As the model data are masked to agree with the observed data coverage, this should solely decrease the power of the detection algorithm by increasing the variance in the dataset, rather than add any systematic bias. The likely impact of this source of uncertainty is indirectly assessed within the present analysis, through changing the number of grid-box values required in any zonal band for the calculation of a zonal mean value. Individual grid-box values are zonally averaged, being weighted to account for changing areal coverage with latitude, to yield a zonal-mean input field to the detection algorithm.

Observed data used in the current study are the HadRT radiosonde temperature dataset (Parker et al., 1997). This is the only available gridded upper air temperature dataset of both suitable length and quality. Previous studies (Barnett et al., 1998 for example), indicate that records should be greater than 30 years in length when used in detection studies, and therefore the MSU temperature record (Christy et al., 2000) is likely to be too short at present. The HadRT dataset is compared to model output from two generations of the Hadley Centre GCM, HadCM2 (Mitchell et al., 1995, Johns et al., 1997), and HadCM3 (Pope et al., 2000, Gordon et al., 2000). HadRT data are available on a monthly basis for the period 1958 to date, on a 5° latitude by

10° longitude grid. HadCM2 and HadCM3 data are available on an annual basis and at a resolution of 2.5° by 3.75°, for the same period. Model data is bilinearly interpolated (Press et al., 1992) to the HadRT grid to enable direct comparisons to be made between the datasets (AT99, Tett et al., 2001). Model data are subsequently spatio-temporally sub-sampled to the available HadRT dataset coverage.

The basic premise of the optimal regression detection approach, detailed in AT99 and AS01 (see chapter 4), is that the observations can be expressed as a linear combination of the model signals, and an additive noise term due to natural internal climate variability. The signals are assumed to exhibit the correct response patterns to the forcings, but not necessarily the correct amplitudes. The problem therefore becomes one of fitting the amplitude, to find the optimal scaling factors of the individual signals, to best reproduce the observations. Optimisation is carried out by rotating the input fields, such that the regression concentrates upon those modes of variability where the signals dominate over natural variability in the model. Because the input data generally have orders of magnitude more data points than the available degrees of freedom of the model control run being used, this optimisation cannot be carried out in full field space. Therefore, the analysis is carried out in the reduced-phase space spanned by the leading EOFs of a section of the model control run. The regression yields a range of plausible solutions that can be explicitly mapped to provide both a best-guess amplitude, and uncertainty limits, for each signal, or any combination of signals. The uncertainty arises from the presence of noise due to natural internal climatic variability both within the observations, and the model fields. It is generally assumed that both observational and model errors are negligible and do not add to the uncertainty in the solution (AS01). In the current study, only uni-dimensional limits are considered as it is aimed solely at addressing simple questions of the detectable presence of individual signals in the observations. The residuals of the regression are compared to an independent section of the control simulation, to provide a check that the regression is a consistent (plausible) explanation of the observations.

Standard tests for degeneracy (Mardia et al., 1979, Tett et al., 1999) indicate that at most three signals can be considered simultaneously in fixed-signal zonal-mean optimal regression detection studies using both HadCM2 and HadCM3. To enable direct comparisons to be made with Tett et al. (2001), the same three-way signal input combination of well-mixed anthropogenic Greenhouse-gases and anthropogenic Sulphate aerosols (GS), stratospheric Ozone depletion (O) and NATural external (solar and volcanic) forcings (NAT) is considered. For HadCM2, the NAT ensemble is calculated by linearly combining the LBB (forced by the Lean et al. (1995) solar forcing reconstruction) and VOL (forced by the Sato et al. (1993) reconstruction of volcanic forcing) ensemble means together. The noise estimate is inflated to account for the additional uncertainty in this composite signal, being the product of two uncertain signals (AT99). These are the two individual forcings combined in the NAT ensemble in HadCM3. There are additional differences between the equivalent forcings (chapter 1, Tett et al., 1999, 2001, Stott et al., 2001), as well as changes between model generations (Pope et al., 2000, Gordon et al., 2000), which provides a degree of independence in the results from the two models. Results of the detection exercise are quoted for two truncations: 21, the estimated degrees of freedom of the available model control data used to perform the optimisation (Allen and Smith, 1997); and 11, half this value, to assess sensitivity to this choice. Where results have been found to be inconsistent when residuals of the regression are compared to an independent section of control, they are clearly labelled as such.

Previous studies have employed a cut-off value of the number of grid-box values required in any latitudinal band for a zonal-mean value to be calculated. In fixed-signal optimal zonal-mean detection studies, three grid-box values within each zone have been required (AT99, Tett et al., 2001). This ensures, at least in part, that the data is representative, although it is by no means perfect as residual observational error could remain in such poorly sampled regions (chapter 2). Such a cut-off is also clearly an arbitrary choice and, therefore, it is of interest to see whether the results of optimal detection studies are sensitive to this choice. In this study the inclusion criteria are incremented by steps of two from one grid box to seven grid boxes required for the calculation of any zonal mean. Expectations are that increasing the coverage requirement will decrease any residual observational error. There are two reasons for this. First, as the number of grid boxes being used to calculate a zonal mean is increased, the true signal will tend to dominate over any residual random errors within the observed dataset. Of course, if there is any residual systematic error in the observations at any given latitude, then this will no longer hold true. Second, in the better-sampled regions, the spatial quality control described in chapter 2 has removed any gross residual errors in the HadRT observations, although smaller errors are likely to remain in both versions. However, increasingly strict inclusion criteria will remove proportionately more data from the poorly sampled tropical and Southern Hemisphere regions than elsewhere and, therefore, there is also a change in geographical emphasis, which is equally likely to affect the results. There is a known tropical upper tropospheric warm bias seen in the models used here when compared to the available observations (Johns et al., 1997, Pope et al., 2000, Gillett et al., 2000). Whether the observational coverage in these regions is adequate, and the data sufficiently accurate, to prove this bias is debatable. Error in the observations masking a true trend agreement with the model predictions in this region can only ever be a theoretical discussion given the available HadRT data.

In optimal zonal-mean detection studies to date the input data have been mass weighted (AT99, Tett et al., 2001, G. S. Jones et al., 2001, Hill et al., 2001). There are at least three plausible candidate weightings that could be applied: “mass weighting”, “volume weighting”, and “no weighting” (Gillett et al., 2000). “Mass weighting” prescribes equal weight to regions of equal mass and, therefore, will tend to focus any statistic upon the troposphere, and particularly the lower troposphere. “Volume weighting” gives equal weight to regions of equal volume and, therefore, emphasises values within the stratosphere. “No weighting” gives each point equal weighting, but it should be noted that, although the levels are separated by approximately equal heights, there is no physical justification for this choice, unlike the other two. The effects of the number of grid boxes in any latitude are not incorporated into any of these weighting schemes, although theoretically such effects could be accounted for in subsequent work. Other sensitivity studies should provide at least some indication as to the likely effects of taking this into account when calculating the weightings. The weighting options are simply linear transformations applied to the raw model and observed zonal-mean data fields. Hence, in the presence of an adequately long control run from which to ascertain an accurate estimate of natural variability, with which to perform the optimisation, the detection results would be identical whichever weighting were applied. However, the available control run contains an order of magnitude fewer degrees of freedom (estimated as 1.5 times the number of non-overlapping chunks, Allen and Smith, 1997) than there are independent datapoints in the zonal-mean difference field being input to the regression. Therefore, expectations must be that the weighting scheme employed will affect the results of any detection study. This study implements all three pre-processing choices to assess the magnitude of this effect.

Previous detection studies considering HadCM2 and HadCM3 upper air temperatures (AT99, Tett et al., 2001, Hill et al., 2001, G. S. Jones et al., 2001) have used the unedited version of HadRT2.1. In the present study, the edited versions (following the quality control detailed in chapter 2) of both HadRT2.1 and HadRT2.1s are employed to gain an estimate of sensitivity to the corrections for suspected inhomogeneities, applied with reference to MSUc data (Christy et al., 1998), following the methodology of Parker et al. (1997). HadRT2.1 has had corrections made throughout its depth, whereas HadRT2.1s has had corrections applied solely within the stratosphere. Comparing results between the two datasets should, therefore, provide an indication as to the likely sensitivity of results to observational uncertainties within the troposphere. A comparison with results from Tett et al. (2001) will further yield a crude estimate of the effect of the removal of obviously dubious values from the observed data (chapter 2).

Three versions of each HadRT dataset (2.1 and 2.1s) are considered, to assess the potential sensitivity to the inclusion criteria used in the calculation of individual annual grid-box observed values. Each version requires at least three observed seasonal mean values in any given year (December to November years), for an annual value to be calculated for any grid-box. Version 1 (V1) requires a single month in any season, Version 2 (V2) two months, and Version 3 (V3) the full three months. As the criteria become increasingly strict, coverage decreases, but this should be balanced against the likely reduction in sampling error in the resulting dataset. In each case the annual model output data are used; ideally seasonal or monthly values would be employed to enable a direct comparison. The magnitude of any model sampling errors is likely to be considerably less than that in the observations, so this should not have a first order effect upon results.

Zonal-mean detection studies have been criticised for reflecting to a large part the recent trend differential between stratospheric and tropospheric temperatures, with a cooling stratosphere and warming troposphere, rather than details of the pattern (Legates and Davies, 1997). However, many of the specific criticisms (Legates and Davies, 1997) have been shown to be based upon the particular crude hypothetical example employed (Wigley et al., 2000). To directly address this concern in a rigorous fashion, a version of zonal means incorporating only those values at or below 300 hPa is considered here, as well as the total column zonal field. This circumvents the criticism by being a solely tropospheric input diagnostic, but at an obvious potential cost if a component of the true anthropogenic zonal-mean temperature signal is a differential temperature trend across the tropopause. A stratospheric series could also be constructed in a similar manner, but data coverage is greatly reduced within the stratosphere and the observational errors may well be non-negligible at these altitudes (Gaffen et al., 2000a). Furthermore, Gillett et al. (2000) have shown that within the tropical stratosphere, HadCM2 is an inadequate representation of the (untreated) HadRT2.1 observations, primarily because it does not contain a representation of the Quasi-Biennial Oscillation (QBO). Collins et al. (2001) state that HadCM3 is also unlikely to adequately capture the observed stratospheric temperature variability. It is unrealistic, on theoretical grounds, to expect the current generation of GCMs to adequately capture stratospheric variability and trends, given their extremely coarse vertical resolution in this region of the atmosphere.

Finally, AT99 and Tett et al. (2001) analyse results using an Ordinary Least Squares (OLS) algorithm (AT99), which does not explicitly account for the presence of noise within the signal estimates due to finite ensemble size. A Total Least Squares (TLS) algorithm now exists to explicitly account for this source of noise in the signals, as well as the observations, in climate change detection studies (AS01). Comparison studies have shown that this has little effect upon the principal results of detection studies considering global near-surface temperatures for HadCM2 (Stott et al., 2001a). Both the TLS and the OLS solutions are considered here to assess the sensitivity of results to taking into account noise in the model signals when considering zonal-mean temperatures. Both have been discussed in detail in chapter 4.

5.2 Assessing the sensitivity of fixed-signal zonal-mean optimal detection results

This section begins by considering the effects of data inclusion criteria, to graphically assess the impact upon both the coverage and temperature field pattern characteristics of the raw input data. The changing observational data mask with increasing criteria of the number of grid boxes required in any given latitude for a zonal mean value to be calculated for V2 of HadRT2.1s is shown in Figure 5.1. In the case of a single value being required, there is data at almost all latitudes and pressure levels, with large areas of missing data only in high southern latitudes. As the coverage criteria become increasingly strict, it can be seen that there is a disproportionate loss of tropical and Southern Hemisphere data. There is also a reduction in the stratospheric data coverage. However, the overall observed temperature field pattern becomes more coherent as the number of required grid boxes for any latitude increases, especially within the stratosphere. This implies that, at least in part, the sampling error of the resulting dataset is being decreased by removing data from these poorly-sampled latitudes. Another view, which cannot be discounted, is that at least some of the higher variance in poorly-sampled regions could be a true manifestation of the climate system. Given the two versions of the HadRT dataset available here, it is not really possible to determine which parts of the observations in poorly-sampled regions are noise due to sampling error, and which are true observed temperatures.

The effects of the different weighting schemes employed on the raw input zonal-mean observed temperature fields are shown in Figure 5.2 for the case where three values are required for the calculation of a zonal mean. For the "mass weighting" scheme, observed gradients in stratospheric temperature trends, especially within the tropics, are greatly reduced. The majority of the pattern detail is within the troposphere, although a large part of the observed pattern arises from the temperature trend differential across the tropopause. For "volume weighting", tropospheric temperature values are heavily damped, but again a large component of the pattern is derived from a divergence in trends across the tropopause. The case of "no weighting" is intermediate between the other two cases, although is clearly closer to the volume weighting, at least in this graphical representation.

Table 5.1 summarises the results of using an OLS approach addressing uncertainties due to weighting algorithm, criteria on the number of reporting grid boxes required for each latitude, version of HadRT data used as the observed input data, version of Hadley Centre model, and truncation. The most immediate conclusion to be drawn from this table, is that the probability of a successful outcome is critically dependent upon the weighting scheme employed. In very few cases for "volume weighting" or "no weighting" are the residuals from the regression consistent. The most likely reason for this is a combination of increasing observational error with altitude (Gaffen et al., 2000a), and gross model deficiencies within the stratosphere (Gillett et al., 2000, Collins et al., 2001). As the truncation increases, the residuals of EOFs with high weighting in the stratosphere are likely to become very large, as EOFs based upon unrealistic model modes of variability are incorporated (see AT99, and

Tett et al., 2001, for a more complete discussion). The "mass weighting" scheme damps down the stratospheric components, concentrating instead upon the troposphere, for which greater confidence exists in both the model and the observations (Gillett et al., 2000, Tett et al., 2001, Gaffen et al., 2000a, Collins et al., 2001). Even when the test on the residuals passes, there is no guarantee that any of the input signals are detected. In the few cases where residuals are consistent in the "volume weighting" and "no weighting" schemes (HadCM3 only), GS is always detected, although its amplitude is found to be significantly overestimated in the model. In the case of "volume weighting", there is also a stratospheric ozone signal detected when using the HadCM3 signals, which is an encouraging result, given the fact that in this case the fields are weighted to emphasise stratospheric values.

From Table 5.1 it can also be seen that the residuals are more likely to be consistent at the lower truncation of 11 than at truncation 21, for all combinations considered. This result is consistent both with the analysis of Tett et al. (2001), although they were not able to consider anything greater than truncation 7 (beyond which in their analysis the residuals became inconsistent), and theoretical expectations. As the truncation is increased the regression will begin to incorporate modes of variability which are poorly sampled in the finite section of the model control run used for optimisation, and these modes will be given artificially high weight in the optimised signals (AT99). The residuals of the regression will become increasingly unrealistic as these modes are introduced, and therefore tests on the residuals will fail (AT99). Tett et al. (2001) conclude that the most likely cause of the consistency failure at relatively low truncation in zonal-mean studies is the presence of gross model inadequacies within the stratosphere, even considering results from a mass weighted scheme. It is worth noting that the results here are consistent with those of Tett et al. (2001), the closest corollary being mass weighting, 3 grid box values for a zonal mean required, and HadCM3 with HadRT2.1, and for which it can be seen from the table the statistical test on the residuals fails at truncation 11. However, in the current analysis, results from this input combination (not shown here) can be considered for truncations up to 9 before the residuals become inconsistent. The major difference between these studies is that the current study uses an updated version of HadRT2.1, in which gross residual errors have been removed (chapter 2). Therefore, weak evidence points to the removal of obviously dubious values from the HadRT dataset, increasing the range of truncations that can be considered. It should also be noted that Tett et al. (2001) were sub-optimal in their choice of pre-processing, and of their model and observational dataset input combination. Other combinations allow for greater truncations, up to at least truncation 21. With increasing truncation, the power of the detection algorithm to differentiate between competing forcing mechanisms is increased. Therefore, Tett et al. (2001) may have reduced by chance their likelihood of making meaningful conclusions about the causes of recently observed trends in zonally-averaged upper air temperatures by sampling from a single input pre-processing combination. Results indicate that the choice of Tett et al. (2001) led to relatively early residuals consistency test failure when compared to the entire range of possible solutions.

Of most interest from Table 5.1 is that results appear to be critically dependent upon the choice of HadRT dataset used as input to the OLS algorithm. Corrections have been applied within the troposphere to HadRT2.1 data using the methodology of Parker et al. (1997), with reference to the MSUc satellite temperature records (Christy et al., 1998, 2000). Only one record (albeit in numerous versions) of this exists, and there are several remaining uncertainties (NRC, 2000). Figure 5.3 shows the two input HadRT fields and the fractional difference between them. Corrections applied within the troposphere are in some regions quite large, being of greater magnitude than the observed trend in some latitudes and at some heights. Many of these large fractional differences tend to occur in regions of small magnitude zonal-mean temperature differences between the periods, and therefore may solely be a statistical artefact due to these small trends in the raw data. It can be inferred that in the majority of cases tropospheric corrections have been relatively minor, although there are likely to exist significant differences between the datasets in at least some regions. It is not possible from a purely visual inspection, to make any meaningful conclusions about which dataset is likely to be a better representation of the “true” observations.

Both versions of HadRT used here are likely to contain at least some residual errors (chapter 2, Gaffen et al., 2000a, Parker et al., 1997). The development of further, independently produced, observational datasets of comparable length is imperative to reduce observational uncertainties further (Santer et al., 1999, NRC, 2000). Combinations with HadRT2.1s as the observed dataset consistently yield more OLS detection results which have residuals that pass the consistency test. This implies that corrections made to HadRT within the troposphere (Parker et al., 1997), result in patterns of temperature changes that are dissimilar to anything seen in the model world for both models, at least in their leading modes of natural variability (see also discussion in section 6.8.1). Further, the corrections could be removing a component of the true signal within the observations. However, the systematic bias in results is almost entirely concerned with whether the residuals are consistent. There is no noticeable systematic effect on which (if any) signals are detected when the residuals are consistent. Therefore, it is unlikely that for this particular zonal-mean difference field, the corrections have removed a significant proportion of the signals being considered. The possibility for a systematic effect of the choice of HadRT dataset version used upon which signals are detected cannot, however, be entirely ruled out for other choices of upper-air temperature input diagnostics based solely upon this analysis.

To a lesser extent, results appear to be model dependent, especially so at higher truncations, when only HadCM2 fields are consistent. This is likely to relate, at least in part, to known natural internal variability underestimation problems within HadCM3 (Collins et al., 2001). Results are also dependent to a slight degree upon the number of values required for inclusion in the zonal average (see Figure 5.1). Regardless of all the considerations, if any signals are detected in any particular OLS analysis in Table 5.1 then GS is always detected, although occasionally it is found that the model GS field is an inconsistent explanation of the observations at the 90%

confidence interval. At small truncations the model tends to significantly overestimate the GS signal strength (yielding an estimated scaling amplitude significantly less than unity), whereas at larger truncations it tends to underestimate the signal. Analysis of a series of best-guess amplitude estimates with increasing truncation indicates that the inclusion of more information has a systematic effect upon the estimated amplitude of the model simulated anthropogenic effect (GS). Figure 5.4 gives an example in which all three input signal best-guess amplitudes exhibit distinct (although discontinuous) positive trends with increasing truncation. Results for other combinations of input fields and pre-processing (not shown) also exhibit very definite increases in the amplitude estimates with truncation, especially for GS; hence the results detailed in this figure are not a statistical sampling artefact. The trend in best-guess amplitude estimates is similar if a TLS algorithm is employed, although for TLS it is much less marked. Analysis of SNRs in Table 5.9 indicates that the SNR for GS is sufficient for the signal not to be significantly noise contaminated for all weighting schemes, and at both the truncations considered. Therefore, the increasing trend in the individual amplitude estimators is unlikely to be due to gaining a stronger (less noise polluted) signal realisation with increasing truncation, which can lead to underestimation problems (AS01). The most likely reason for any residual (true) increasing trend is that at low truncations the input to the regression is likely to be dominated by the global mean change, which in both models is greater in their GS runs than the observations, particularly for HadCM3 (chapter 3). As truncation increases, regional patterns of change will tend to dominate the signal, and these regional gradients may well be underestimated in the models. The potential for systematic non-stationarity of regression amplitude estimates must be borne in mind when potential applications of the detection algorithm (Allen et al., 2000, 2001, for example) are being considered. Both stratospheric ozone, and natural external forcings (solar and volcanic effects), are detected in Table 5.1 for some combinations of pre-processing and input choices, but are much more sensitive, and therefore there is lower confidence in their presence in the observations than there is for a GS signal.

Employing a TLS algorithm to take noise into account in the model signal response estimates greatly impacts the results (Table 5.2, c.f. Table 5.1). The most immediately obvious impact of taking noise into account in the model signals, is that there are far fewer cases where the residuals are inconsistent. This is encouraging as, in taking this additional noise term into account in the analysis, expectations must be that the residuals will more closely match an independent estimate of the noise due to natural variability. “Mass weighting” is found to be the only scheme for which consistent residuals are readily attainable, in common with the OLS approach, although again this does not guarantee the detection of any of the input signals. It is only at the higher truncation of 21 that signals become detectable, although there is a systematic underestimation problem in the model signals at these truncations, noticeably more so than under the OLS approach. This is consistent with theoretical expectations that, for small ensemble sizes, OLS will tend to underestimate the true signal amplitude and, therefore, overestimate the signal strength in the observations (AS01, Stott et al., 2001a). The fact that at truncation 11 the OLS approach robustly yields a detectable GS signal, whereas the TLS approach does not, would appear to reduce confidence in the results for both approaches. By incorporating noise from the signal estimates into the detection algorithm, the resulting GS signal strength uncertainty range is no longer significantly different from zero. Analysis of the raw results shows that the difference between the two approaches is that the TLS uncertainty range is much greater, rather than there being any fundamental differences in the best-guess GS signal amplitudes between the approaches. AS01 note how the presence of non-negligible observational error could lead to problems in employing a TLS approach, although if its covariance structure were known it could theoretically be incorporated into the algorithm (AS01). The likelihood is that with a zonal-mean input dataset the observational error is non-negligible (Parker and Cox, 1995, Gaffen et al., 2000a, amongst others) and, therefore, it is possible that the uncertainty estimates in a TLS approach could be overestimated, particularly at these small truncations. However this has not been specifically tested. Results considering a TLS approach are less dependent upon the version of HadRT or Hadley Centre model used as input to the detection algorithm than is the case for OLS, although there is still a degree of sensitivity to both.

In all but one of the cases where a signal, or combination of signals, is detected using a TLS approach, a NAT signal is found, with a GS signal also being detectable in the majority of cases. The presence of a detectable O signal is far less certain. The main difference in terms of signals detected in employing the TLS algorithm is, therefore, that the relatively weak NAT signal has become more detectable, but evidence for a tropospheric anthropogenic effect (GS) remains strong, although not unequivocal. Analysis of SNR values for both models in Table 6.9 shows that NAT is consistently the weakest signal and that, for HadCM3 "mass weighted" fields, it is significantly noise contaminated. Therefore, the lack of positive OLS detections may solely relate to known negative biases in the OLS estimator for weak signals (AS01), particularly for "mass weighted" input. In an attempt to confirm these findings, the input (pre-whitened (optimised)) ensemble mean fields, and observations, are shown in Figure 5.5 for HadCM3. The natural forcings integration has the smallest amplitude of all these ensemble means, and yet in this three-way TLS regression it is the most robustly detected. Precise reasons for this result are unknown, as by a purely visual comparison the GS signal most resembles the observations. Natural forcings may modify the pattern of the GS response to better match the observations, and hence be detected. GS is overestimated in amplitude in the model and, therefore, its TLS amplitude estimate may be indistinguishable from zero in some cases, leading to occasional non-detection. The patterns shown in Figure 5.5 are qualitatively similar at truncation 21 and, therefore, results are likely to be similar over the range of truncations between the two truncations that were considered in the present study.

Up until now, results discussed have been those using V2 of the HadRT datasets as the observational input fields to the regression. In Tables 5.3 and 5.4, the analysis is repeated for mass weighted input fields considering all three versions of each HadRT dataset. Table 5.3 shows that when considering results from an OLS approach, the residuals tend to become more consistent with increasingly strict inclusion criteria for individual grid-box values (V1 to V3). That is, in an OLS approach, decreasing the potential sampling error for the individual grid boxes means the residuals (which are likely to include at least a proportion of this error) become more similar to an independent estimate of the natural variability. There appears to be no cost function (in terms of the power of the algorithm) involved in decreasing the coverage in this manner, as all signals remain detectable. In all cases where signals are detected, GS is found, so this result is not sensitive to this inclusion criteria. TLS results shown in Table 5.4 exhibit much lower variability, at high truncations, than the OLS results shown in Table 5.3; they are more stable to individual grid-box sampling uncertainty at higher truncations. The sampling affects, to a lesser degree, the simulated signals as well as the observations, and the TLS algorithm accounts for noise in the signals, whereas the OLS algorithm does not. This is therefore an encouraging result. At the lower truncation of 11 considered here the TLS results are critically dependent upon the version of the observed and modelled datasets being considered. For V1, there is a strong result that a natural forcings signal is detectable for HadCM2 with HadRT2.1s, but this is not evident in versions V2 or V3, or for other model / observed dataset combinations. The implication is that the use of single datasets without explicitly dealing with potential sources of uncertainty could be misleading in detection studies. The systematic difference in results at truncation 11, noted previously, between the two regression approaches is insensitive to the choice of version of the HadRT observational dataset used.

Finally, addressing concerns that results of previous zonal-mean detection studies are critically dependent upon the presence of the troposphere-stratosphere divergence in temperature trends (Legates and Davies, 1997), the analyses in Tables 5.1 to 5.4 are repeated in Tables 5.5 to 5.8 for a troposphere-only zonally-averaged input field. Table 5.5 (c.f. Table 5.1) shows that OLS results are no longer as sensitive to the weighting applied as is the case when considering the total column zonal mean field. Differences between the weighted fields are reduced in this diagnostic (see the portions below 300 hPa of the fields in Figure 5.2), as the levels are approximately equi-distant, and contain similar proportions of the total atmospheric mass.

Therefore, expectations are that results will be less dependent upon the pre-weighting applied to the fields. Other conclusions remain the same as those for a total column zonal mean field, although a stratospheric ozone signal becomes a lot harder to detect. For the TLS approach (Table 5.6), results are again less dependent upon the weighting applied, but otherwise are broadly consistent with those for the full zonal-mean field. It is worth noting, that in some cases employing both regression approaches, a stratospheric ozone signal is still detected, despite the input diagnostic being solely tropospheric. Therefore, at least some of the stratospheric ozone depletion signal has a realisation within the troposphere, although it is implied to be very weak, as results are highly sensitive to methodological uncertainties. Tables 5.7 and 5.8 show that, as is the case for the full zonal-mean difference field detailed previously there is some dependency upon the criteria for the calculation of individual grid-box annual values. In removing the stratospheric data, the seemingly anomalous detection of a natural forcings influence using TLS for V1 of HadRT2.1s with HadCM2 at truncation 11 disappears, reducing confidence in this result further.

5.3 Conclusions

Previous zonal-mean temperature climate change detection studies have found a detectable anthropogenic influence and, more tentatively, natural (solar and volcanic) influences (Santer et al., 1996a, Tett et al., 1996, 2001, AT99, Hill et al., 2001, G.S. Jones et al., 2001). This study has employed variations of the methodology of Tett et al. (2001) and AT99, to assess the sensitivity of these results to at least some of the major likely causes of uncertainty. Here, the HadRT radiosonde temperature record (Parker et al., 1997) has been compared to output from both HadCM2 (Johns et al., 1997) and HadCM3 (Pope et al., 2000). The primary conclusion is that a detectable signal of combined well-mixed anthropogenic greenhouse-gases and sulphate aerosols (GS) is found to be relatively robust to numerous uncertainties, giving increased confidence in an anthropogenic cause for at least part of the recently observed changes in zonally averaged temperatures. There exists, however, a huge range of potential solutions, and in the majority of cases considered in this study, either no signals are detected at all, or the residuals are inconsistent. The use of a single choice of input fields and pre-processing combinations could, therefore, yield highly sub-optimal results. However, in no combinations does this study yield spurious outlier results, which would lead to completely incorrect statements being made as to the causes of recently observed zonally mean temperature changes. Hence, results from previous studies can only ever have been conservative, quoting no detectable influence of a particular forcing when such a forcing were in fact detectable. In some cases, the GS signal of the models is found to be an inconsistent explanation of the observed changes at the 90% confidence interval, reducing confidence in the ability of the models. Detection of natural external forcings and a stratospheric ozone depletion signal are found to be less robust to uncertainties, although this does not rule out their presence in the observations.

Under both an OLS, and to a lesser extent, a TLS approach, there is a systematic trend of GS amplitude estimates with increasing truncation (the number of principal modes of variation being considered). According to the results considered here, at low truncations the models tend to overestimate the GS signal strength, whereas at high truncations they tend to underestimate the response. This is unlikely to be a statistical artefact, as at low truncations the SNR is not low enough to yield negatively biased estimators for the OLS approach (AS01, Tett et al., 2001). Additionally, the TLS approach also exhibits trends, so at least part of these trends are likely to be real, although the reasons are unknown. Regardless of the reasons for these observed trends, the potential for significant non-stationarity of signal amplitude estimates with truncation should be explicitly considered when applying the detection algorithm to bounding uncertainties in projections of future climate (Allen et al., 2000, 2001).

Implementing a “mass weighting” scheme rather than “volume weighting” or “no weighting” on the input data, improves the chances of a successful detection exercise. It is believed that this primarily reflects the uncertainties in both the models and the observations within the stratosphere (Gaffen et al., 2000a, Gillett et al., 2000, Collins et al., 2001), a region which will be given low weighting when using a "mass weighting" scheme. Equally, consideration of a troposphere-only signal, which removes the troposphere-stratosphere temperature trend difference, does not affect the primary conclusions, although a stratospheric ozone depletion signal becomes much less detectable as there is no stratospheric field. Importantly, this shows that detection of a GS signal is not critically dependent upon the presence of the divergence in temperature trends across the tropopause in both the signals and the observations, as has been previously suggested (Legates and Davies, 1997). The observed tropospheric warming is outside the control range for both models according to the results detailed here.

Varying criteria for data inclusion for the calculation of the HadRT observational dataset grid-box annual values, and the number of those grid boxes required in any latitudinal band to enable a zonal mean value to be calculated, have been found to have an effect upon the results, particularly when using OLS regression. The effects of sampling error and geographical coverage changes cannot be truly separated under the approach detailed here. There is no discernible systematic bias in the signals being detected with the changing criteria. However, the truncation at which the residuals become inconsistent is seen to vary widely based solely upon these criteria. By chance, Tett et al. (2001) chose a pre-processing combination that yielded inconsistent residuals at the lower end of this truncation range, which may have artificially limited the power of their analysis. By choosing an alternative input pre-processing combination the truncation could have been increased giving more power to their detection algorithm.

Whether HadRT2.1 or HadRT2.1s are used as the observational dataset has an impact upon the results using both regression approaches. Use of the troposphere-corrected HadRT2.1 dataset significantly reduces the number of cases with consistent residuals, particularly for an OLS regression approach. The implication of this is that corrections applied to the HadRT dataset by Parker et al. (1997) force it to exhibit modes of variability not seen in the model world (at least in their leading modes). It is tentatively suggested that this behaviour could also be due, at least in part, to corrections being made which remove a component of the true signals from the observations, although there is no robust evidence for this. Use of a single observed dataset, without accounting for likely errors and uncertainties in a rigorous manner could, therefore, lead to sub-optimal conclusions. Importantly, there is no systematic bias in which signals are detected and, therefore, it is concluded that not accounting for potential observational uncertainties does not, at least in the case of zonal-mean fields, yield spurious detection results.

Accounting for noise in the signals by employing a TLS approach tends to increase the chances of detecting both natural and stratospheric ozone forcings, and increase the number of EOFs which can be considered in the regression. However, at the lower truncation considered here, the OLS approach yields a detectable GS signal, whereas the TLS approach does not. This primarily reflects an increase in the uncertainty estimate under TLS, rather than any systematic bias in the best-guess amplitude estimates. This increased uncertainty is most likely to be due, at least in part, to the presence of non-negligible residual observational errors within the HadRT dataset. Ideally these would be included within the analysis, but there is no obvious way of accurately deriving an unbiased estimate of the observational error covariance matrix (AS01).

Finally, the detection of a GS signal has been found to be relatively insensitive to choice of Hadley Centre GCM. Although HadCM2 and HadCM3 are based on the same underlying physical model, there are sufficient differences between the models (Pope et al., 2000) and forcings (Stott et al., 2001, Tett et al., 2001) to justify treating them as independent simulations. Future work should address results from different, truly independent, models and observed datasets, as far as such independent datasets are possible, to confirm these conclusions.

Results of OLS regression using V2 of HadRT datasets

Truncation 11

Model and HadRT versions	Mass Weighting				Volume Weighting				No Weighting			
	1	3	5	7	1	3	5	7	1	3	5	7
HadCM2 + HadRT2.1s	GS	GS	GS	GS	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
HadCM3 + HadRT2.1s	GS$	GS$	GS	GS	🟥	🟥	GS$,O	🟥	GS$	🟥	🟥	🟥
HadCM2 + HadRT2.1	🟥				🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
HadCM3 + HadRT2.1	GS$,NAT	🟥	🟥	🟥	🟥	🟥	GS$,O	🟥	🟥	🟥	🟥	🟥

Truncation 21

Model and HadRT versions	Mass Weighting				Volume Weighting				No Weighting			
	1	3	5	7	1	3	5	7	1	3	5	7
HadCM2 + HadRT2.1s	GS	GS*,O,NAT	GS*,O,NAT	GS*,O$,NAT	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
HadCM3 + HadRT2.1s	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
HadCM2 + HadRT2.1	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
HadCM3 + HadRT2.1	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥

🟥 Residuals inconsistent

GS* Signal significantly under-estimated

GS$ Signal significantly over-estimated

Table 5.1. Results of an Ordinary Least Squares detection regression algorithm on zonally averaged upper-air temperatures. If a signal is detected then it is quoted in the relevant box. Inconsistent residuals are highlighted. Three pre-processing weightings have been applied to both the observations and the model fields before input to the regression algorithm. For each of these fields the number of grid-box values required for a zonal mean to be calculated has been incremented in steps of two from one to seven. Both HadCM2 and HadCM3 are used to gain an estimate as to the sensitivity to the choice of model signals. HadRT2.1 has had corrections applied with reference to co-located MSUc records within the troposphere (Parker et al., 1997), whereas HadRT 2.1s is the raw observed tropospheric data. Both have been corrected within the stratosphere, and had obviously dubious values removed (chapter 2).

Results of TLS regression using V2 of HadRT datasets

Truncation 11

Model and HadRT versions	Mass Weighting				Volume Weighting				No Weighting			
	1	3	5	7	1	3	5	7	1	3	5	7
HadCM2 + HadRT2.1s												
HadCM3 + HadRT2.1s	NAT				🟥	🟥		🟥	GS$			
HadCM2 + HadRT2.1												
HadCM3 + HadRT2.1	NAT*				🟥	🟥		🟥				

Truncation 21

Model and HadRT versions	Mass Weighting				Volume Weighting				No Weighting			
	1	3	5	7	1	3	5	7	1	3	5	7
HadCM2 + HadRT2.1s	GS,NAT	GS*,NAT*	GS*,NAT*	GS*,NAT*								
HadCM3 + HadRT2.1s				GS*,O*,NAT*	🟥	🟥		🟥	🟥	🟥		
HadCM2 + HadRT2.1		GS,NAT*	GS*,NAT*	GS*,NAT*								
HadCM3 + HadRT2.1			O*,NAT*	GS,O*,NAT*	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥

🟥 Residuals inconsistent

GS* Signal significantly under-estimated
GS$ Signal significantly over-estimated

Table 5.2. As Table 5.1, except employing a Total Least Squares regression algorithm, which accounts for noise in the signals as well as the observations.

Results of OLS regression using mass weighted HadRT datasets

Truncation 11

Model and HadRT versions	**V1**				**V2**				**V3**			
	1	**3**	**5**	**7**	**1**	**3**	**5**	**7**	**1**	**3**	**5**	**7**
HadCM2 + HadRT2.1s	GS	GS	GS	GS	GS	GS	GS	GS		GS	GS	GS
HadCM3 + HadRT2.1s	GS$,NAT	🟥	🟥	🟥	GS$	GS$	GS	GS	GS$	GS$	GS$	GS
HadCM2 + HadRT2.1		GS	GS	GS	🟥				🟥		GS	GS
HadCM3 + HadRT2.1	🟥	🟥	🟥	🟥	GS$,NAT	🟥	🟥	🟥	GS$	GS$,NAT	GS,NAT	GS,NAT

Truncation 21

Model and HadRT versions	**V1**				**V2**				**V3**			
	1	**3**	**5**	**7**	**1**	**3**	**5**	**7**	**1**	**3**	**5**	**7**
HadCM2 + HadRT2.1s	GS	GS*,O,NAT	GS*,O,NAT	GS*,O$,NAT	GS	GS*,O,NAT	GS*,O,NAT	GS*,O,NAT	GS	GS,O,NAT	GS,O	GS*,O$,NAT
HadCM3 + HadRT2.1s	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	GS$	GS	GS,O	🟥
HadCM2 + HadRT2.1	🟥	GS,O,NAT	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
HadCM3 + HadRT2.1	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥

🟥 Residuals inconsistent
GS* Signal significantly under-estimated
GS$ Signal significantly over-estimated

Table 5.3. As Table 5.1, except considering the sensitivity to inclusion criteria for the calculation of individual grid-box annual means. The "mass weighting" scheme is considered, as this yielded the most useful information in Table 5.1.

Results of TLS regression using mass weighted HadRT datasets

Truncation 11

Model and HadRT versions	V1				V2				V3			
	1	3	5	7	1	3	5	7	1	3	5	7
HadCM2 + HadRT2.1s	NAT	GS,NAT	NAT	GS,NAT								
HadCM3 + HadRT2.1s	NAT				NAT							
HadCM2 + HadRT2.1												
HadCM3 + HadRT2.1	NAT				NAT*							NAT

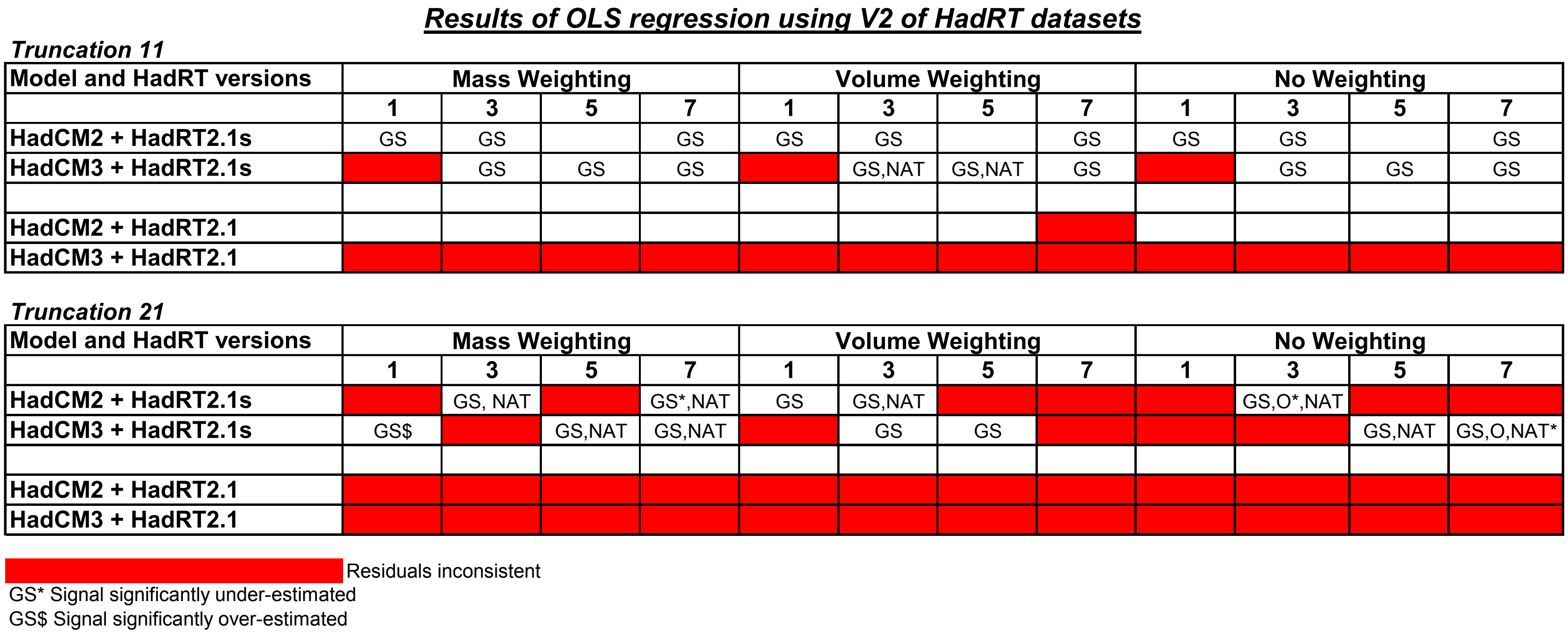
Truncation 21

Model and HadRT versions	V1				V2				V3			
HadCM2 + HadRT2.1s	GS,NAT	GS*,NAT	GS*,NAT*	GS*,NAT*	GS,NAT	GS*,NAT*	GS*,NAT*	GS*,NAT*	GS,NAT	GS,NAT	GS*,O,NAT*	GS*,NAT*
HadCM3 + HadRT2.1s								GS*,O*,NAT*				
HadCM2 + HadRT2.1		GS,NAT	GS,NAT*	GS*,NAT*		GS,NAT*	GS*,NAT*	GS*,NAT*		GS,NAT*	GS*,O,NAT*	GS*,O,NAT*
HadCM3 + HadRT2.1			GS,O*,NAT				O*,NAT*	GS,O*,NAT*			O,NAT*	

🟥 Residuals inconsistent

GS* Signal significantly under-estimated
GS$ Signal significantly over-estimated

Table 5.4. As Table 5.3, except considering the results from a Total Least Squares regression methodology.



Results of OLS regression using V2 of HadRT datasets

Truncation 11

Model and HadRT versions	Mass Weighting				Volume Weighting				No Weighting			
	1	3	5	7	1	3	5	7	1	3	5	7
HadCM2 + HadRT2.1s	GS	GS		GS	GS	GS		GS	GS	GS		GS
HadCM3 + HadRT2.1s	🟥	GS	GS	GS	🟥	GS,NAT	GS,NAT	GS	🟥	GS	GS	GS
HadCM2 + HadRT2.1								🟥				
HadCM3 + HadRT2.1	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥

Truncation 21

Model and HadRT versions	Mass Weighting				Volume Weighting				No Weighting			
	1	3	5	7	1	3	5	7	1	3	5	7
HadCM2 + HadRT2.1s	🟥	GS, NAT	🟥	GS*,NAT	GS	GS,NAT	🟥	🟥	🟥	GS,O*,NAT	🟥	🟥
HadCM3 + HadRT2.1s	GS$	🟥	GS,NAT	GS,NAT	🟥	GS	GS	🟥	🟥	🟥	GS,NAT	GS,O,NAT*
HadCM2 + HadRT2.1	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
HadCM3 + HadRT2.1	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥

🟥 Residuals inconsistent

GS* Signal significantly under-estimated

GS$ Signal significantly over-estimated

Table 5.5. As Table 5.1., except considering a troposphere-only diagnostic. The differences in results between the weighting schemes are much less when compared to those seen in Table 5.1.

Results of TLS regression using V2 of HadRT datasets

Truncation 11

Model and HadRT versions	Mass Weighting				Volume Weighting				No Weighting			
	1	3	5	7	1	3	5	7	1	3	5	7
HadCM2 + HadRT2.1s												
HadCM3 + HadRT2.1s												
HadCM2 + HadRT2.1												
HadCM3 + HadRT2.1												

Truncation 21

Model and HadRT versions	Mass Weighting				Volume Weighting				No Weighting			
	1	3	5	7	1	3	5	7	1	3	5	7
HadCM2 + HadRT2.1s		GS*+O+NAT*		GS,NAT*		NAT		GS*,NAT*		GS*,O,NAT*		GS,NAT*
HadCM3 + HadRT2.1s												GS,O,NAT*
HadCM2 + HadRT2.1				NAT*		NAT*		GS*,NAT*		NAT*		NAT*
HadCM3 + HadRT2.1				GS,O,NAT								GS,O,NAT*

🟥 Residuals inconsistent
GS* Signal significantly under-estimated
GS$ Signal significantly over-estimated

Table 5.6. As Table 5.2, except for a troposphere-only input diagnostic.

Results of OLS regression using mass weighted HadRT datasets

Truncation 11

Model and HadRT versions	**V1**				**V2**				**V3**			
	1	**3**	**5**	**7**	**1**	**3**	**5**	**7**	**1**	**3**	**5**	**7**
HadCM2 + HadRT2.1s	GS	GS*	GS	GS	GS	GS		GS		GS		GS
HadCM3 + HadRT2.1s	🟥	GS	🟥	GS	🟥	GS	GS	GS	GS	GS	GS	GS
HadCM2 + HadRT2.1		GS	🟥	🟥		🟥				🟥		🟥
HadCM3 + HadRT2.1	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	GS	GS,NAT	GS,NAT	GS,NAT*

Truncation 21

Model and HadRT versions	**V1**				**V2**				**V3**			
	1	**3**	**5**	**7**	**1**	**3**	**5**	**7**	**1**	**3**	**5**	**7**
HadCM2 + HadRT2.1s	GS,NAT	🟥	GS,NAT	🟥	🟥	GS,NAT	🟥	GS*,NAT	GS	GS,NAT	GS	GS,NAT
HadCM3 + HadRT2.1s	🟥	🟥	GS	GS	GS$	🟥	GS,NAT	GS,NAT	GS	GS	GS	GS
HadCM2 + HadRT2.1	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	GS,O,NAT	GS,NAT
HadCM3 + HadRT2.1	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥

🟥 Residuals inconsistent
GS* Signal significantly under-estimated
GS$ Signal significantly over-estimated

Table 5.7. As Table 5.3, except for a troposphere-only input diagnostic.

Results of TLS regression using mass weighted HadRT datasets

Truncation 11

Model and HadRT versions	V1				V2				V3			
	1	3	5	7	1	3	5	7	1	3	5	7
HadCM2 + HadRT2.1s												
HadCM3 + HadRT2.1s												
HadCM2 + HadRT2.1												
HadCM3 + HadRT2.1												

Truncation 21

Model and HadRT versions	V1				V2				V3			
	1	3	5	7	1	3	5	7	1	3	5	7
HadCM2 + HadRT2.1s			NAT*	GS,NAT*		GS*,O,NAT*		GS,NAT*		GS,NAT*	NAT	GS,NAT
HadCM3 + HadRT2.1s												
HadCM2 + HadRT2.1			NAT*	NAT*				NAT*			GS,NAT*	GS,NAT*
HadCM3 + HadRT2.1				GS,O*,NAT*		GS,O,NAT						

🟥 Residuals inconsistent

GS* Signal significantly under-estimated

GS$ Signal significantly over-estimated

Table 5.8. As Table 5.4 for a troposphere-only input diagnostic.

	HadCM2			HadCM3		
Weighting and truncation	GS	O	NAT	GS	O	NAT
Mass weighting, 11	9.15	3.38	2.35	17.00	3.25	*1.10*
Mass weighting, 21	6.25	4.32	2.40	10.42	4.03	*1.24*
Volume weighting, 11	11.71	23.90	2.23	17.94	65.33	5.06
Volume weighting, 21	8.97	38.45	2.96	10.76	69.83	6.85
No weighting, 11	11.95	15.00	2.88	16.22	21.32	2.65
No weighting, 21	9.63	35.66	2.72	12.06	44.29	4.31

Table 5.9 SNRs for the different input combinations considered for V2 field coverage. Those cases where they are indistinguishable from noise at the 90% confidence interval are shown in italics.

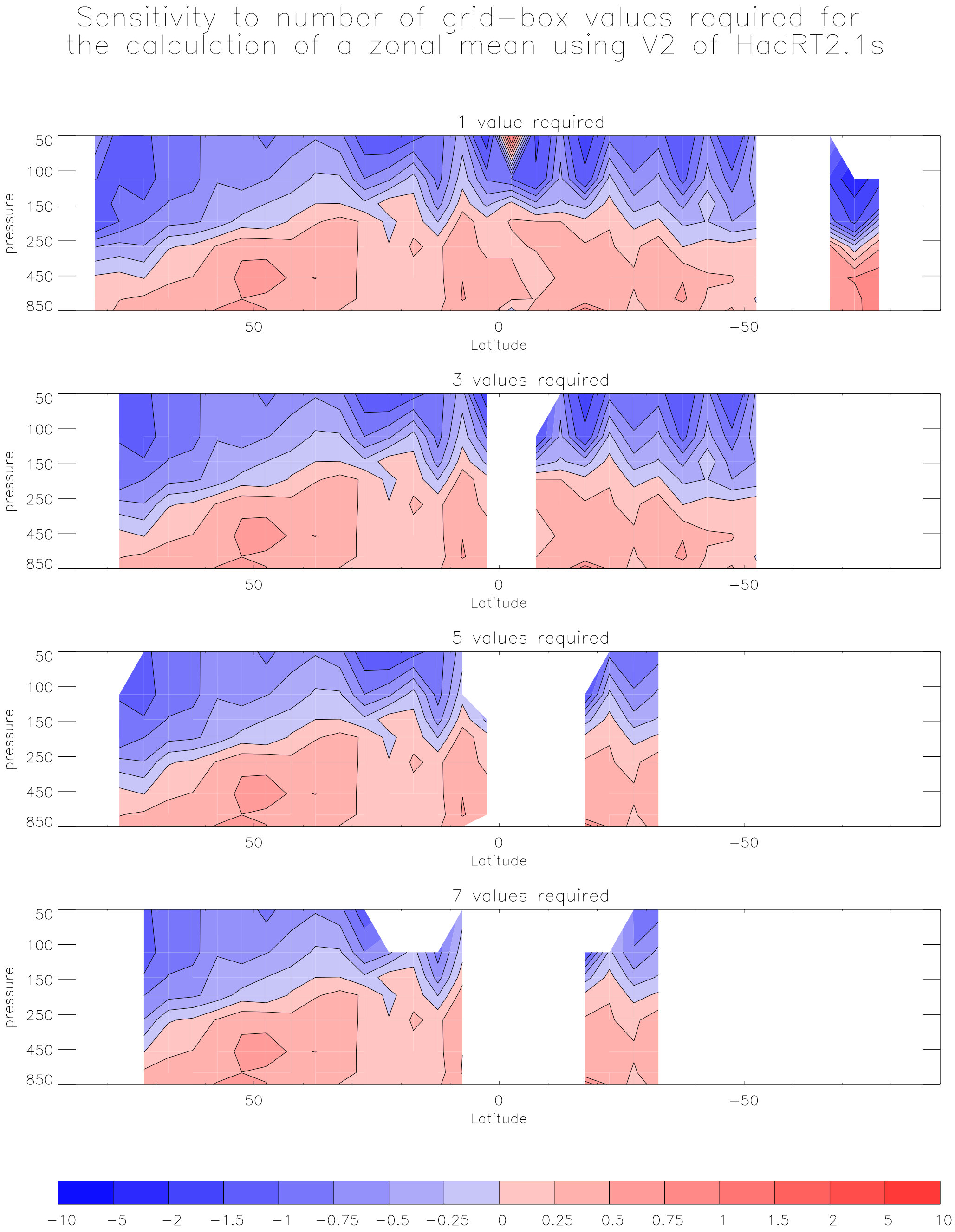


Figure 5.1. Changing coverage characteristics and temperatures for V2 of HadRT2.1s with increasingly strict inclusion criteria for the calculation of a zonal mean value.

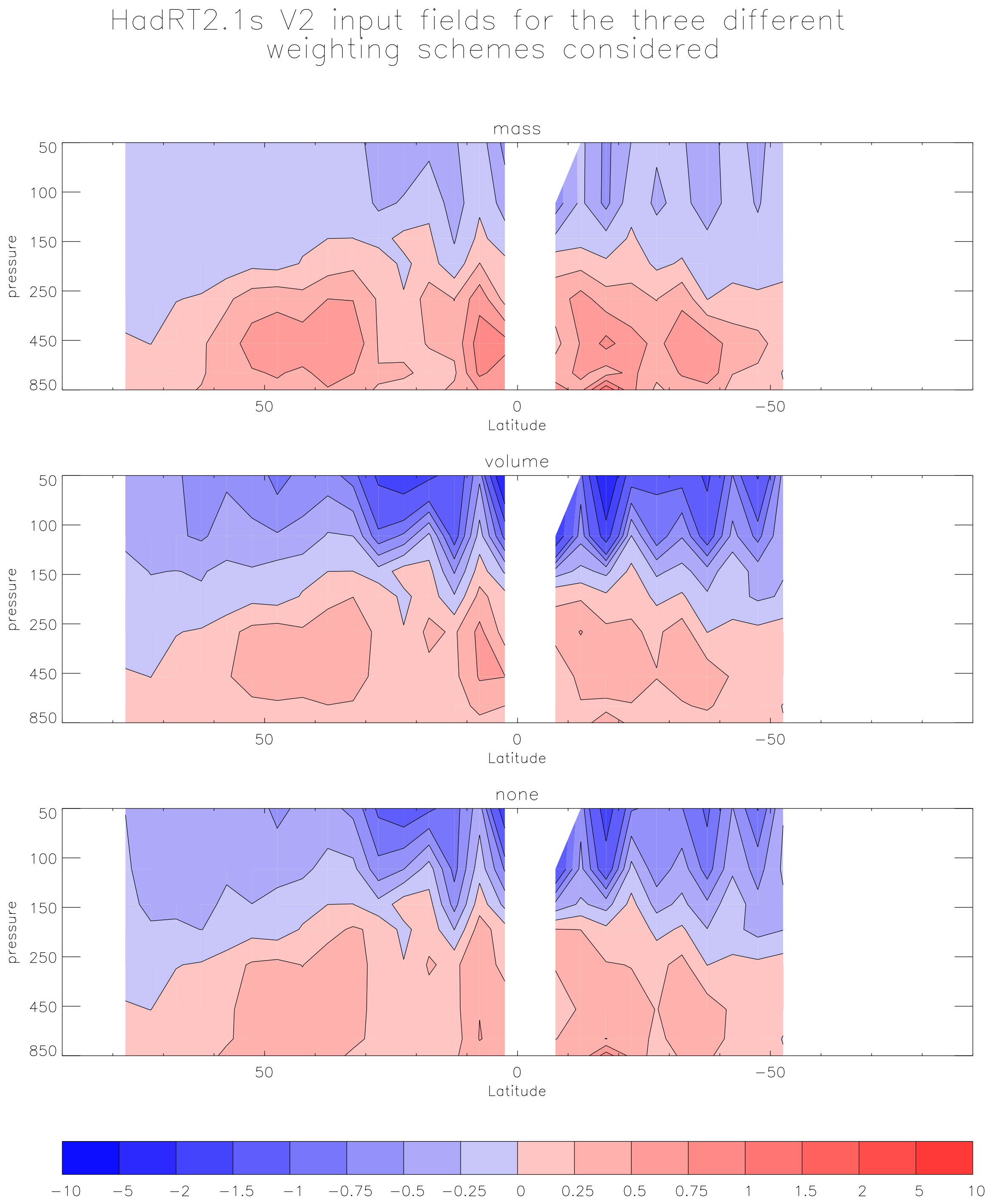


Figure 5.2. The effects of implementing the different weighting schemes on zonal-mean temperatures for V2 of the HadRT2.1s observations. The resulting patterns have been normalised to enable a meaningful visual comparison to be made.

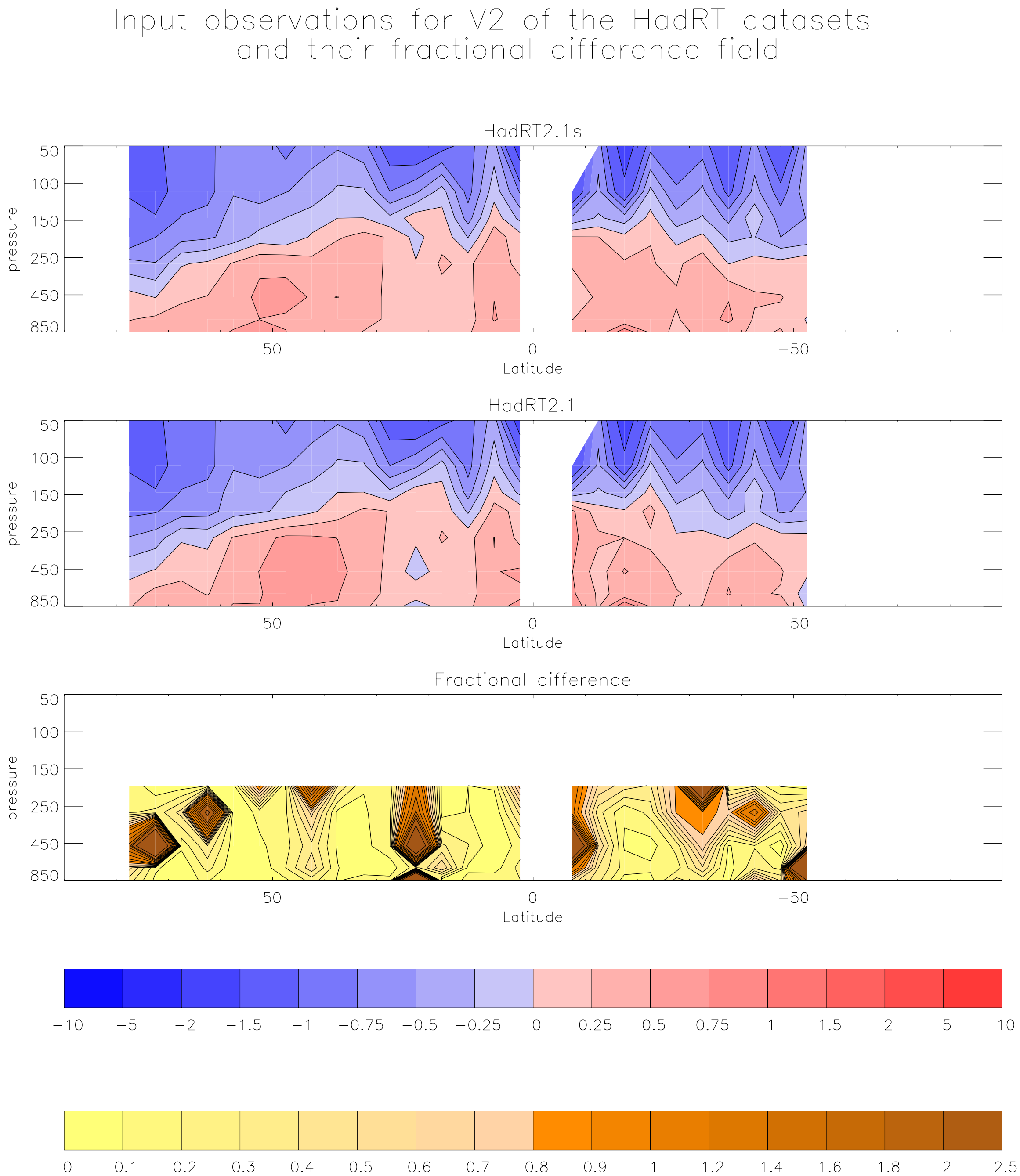


Figure 5.3. Zonally averaged V2 HadRT2.1 and HadRT2.1s temperature data and their fractional difference field. Areas of large fractional differences are usually close to the tropopause. Noticeable differences occur at 25°N and 10°S within the troposphere, and in no regions is there absolutely no change between the series.

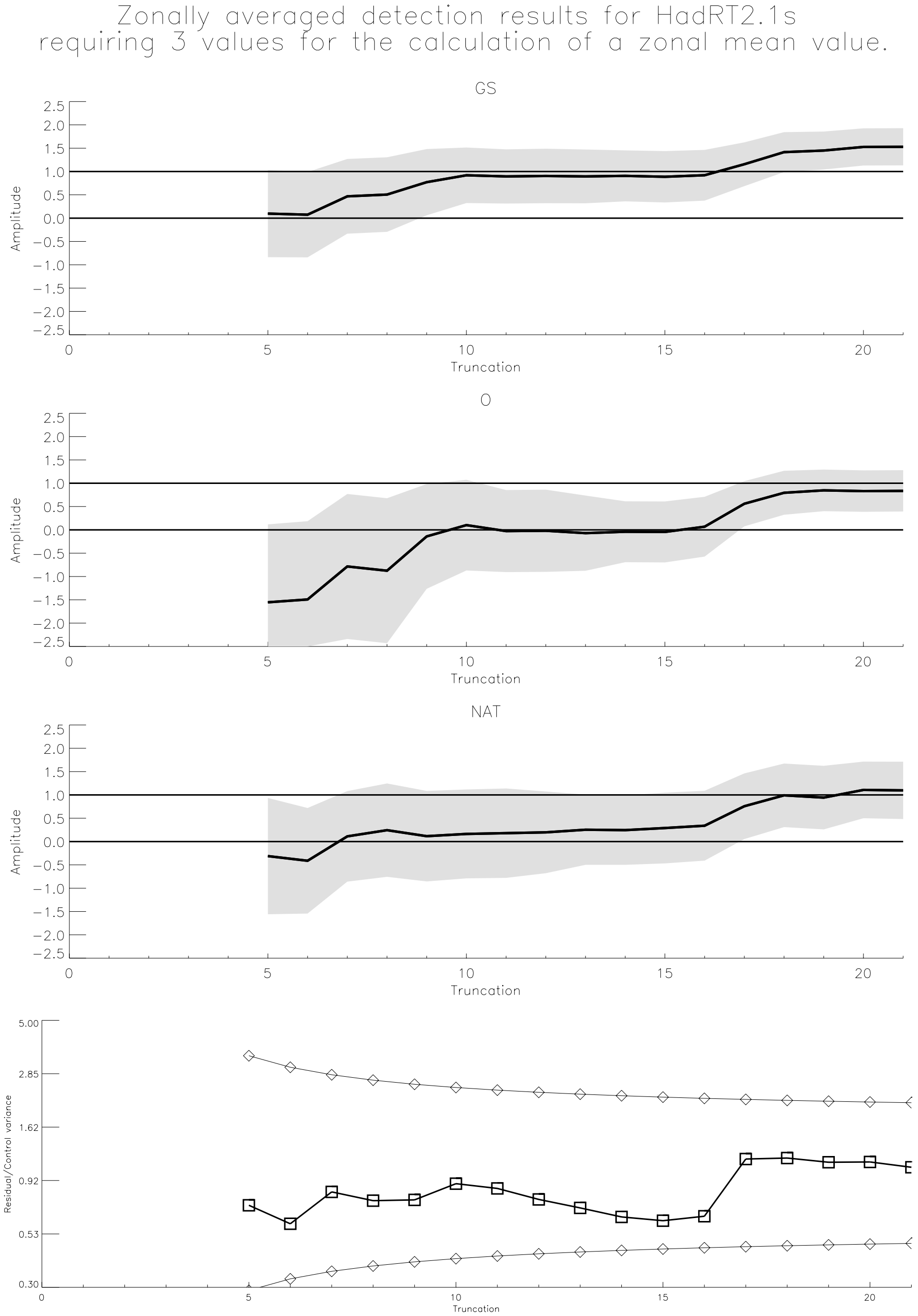


Figure 5.4. Trends in the individual best guess, and range of signal strengths in the observations for increasing truncation. The bottom panel depicts the residuals (square points), which are consistent at all truncations (fall within the F-distribution (diamond points)) for this input combination.

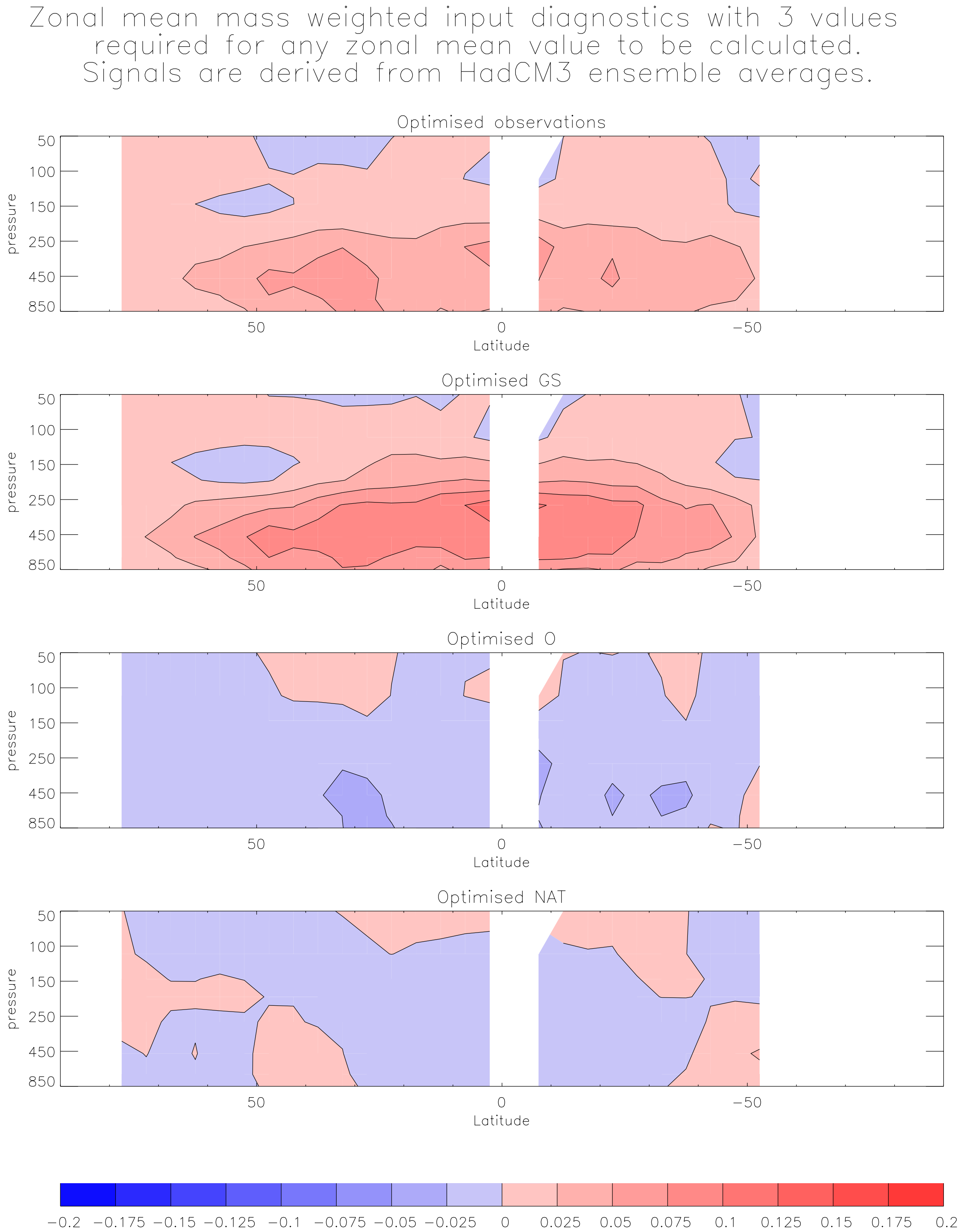


Figure 5.5. Optimised input fields (at truncation 11) to the detection algorithm for "mass weighted" HadCM3 signals. The observed data field is HadRT2.1s.