

List of Figures

Figure 1.1	Global and Hemispheric mean temperatures from the Jones et al. (2001) dataset. Reproduced from http://www.cru.uea.ac.uk/cru/data/temperature/	32
Figure 2.1	Global and hemispheric annual average temperatures for the HadRT dataset at the 500hPa level using the three inclusion criteria. V1 is denoted by the black lines, V2 (the default version) green lines, and V3 red lines. Where values are identical they are overlain. A global coverage statistic is also shown to denote the effects on the spatial completeness of each record. Grid boxes are only included in this analysis if they contain primary observational data.	54
Figure 2.2	The effects of removing infilled grid box values from V2 of the HadRT2.1s data. Coverage is greatly reduced, but there is little difference in the global and hemispheric means. For comparison purposes the green line in Figure 2.1 is equivalent to (but not directly the same as) the red line in this Figure.	55
Figure 2.3	Grid box errors summed over all pressure levels, different error types have been given different values, based upon their perceived importance. Individual Z-Scores have a value of 1, average Z-Scores 100, and correlations 1000. These were chosen so that different error types would be clear in this graphical representation.	56
Figure 2.4	Global field of grid box standard deviations at the 850hPa level. The field is seen to behave in a relatively smooth and coherent manner, although two anomalous points stand out, one in South Western North America and one in Southern Africa.	56
Figure 2.5	Grid box temperature series on levels for the South East Asian grid box. Observations are represented by crosses, with shading to $\pm 2\sigma$; near neighbour averages are represented by diamonds. Note that it is the y values that appear dubious, especially at height.	57
Figure 2.6	Grid box temperature series on levels for an example Indian grid box. Observations are represented by crosses, with shading at $\pm 2\sigma$; near neighbour averages are represented by diamonds. Note significant breakpoints in 1969 and 1990.	58

Figure 2.7	Grid box temperature series on levels for the South African grid box. Observations are represented by crosses, with shading to $\pm 2\sigma$; near neighbour averages are represented by diamonds. Note significant anomaly at the 850hPa level.	59
Figure 2.8	Grid box temperature series on levels for the New Caledonian grid box. Observations are represented by crosses, with shading to $\pm 2\sigma$; near neighbour averages are represented by diamonds. Note significant warm anomaly at height in the early record.	60
Figure 2.9	Plot showing the effect of editing the HadRT dataset, based upon quality control checks, on global and hemispheric mean temperatures and coverage at the 500hPa level. The unedited version is shown as a black line and is the same as the red line in Figure 2.2; the red line is directly analogous to the green line in Figure 2.1 and shows the effects of editing.	61
Figure 2.10(i)	Decadally averaged zonal mean temperatures from V2 of the HadRT 2.1s data as anomalies from 1971-90 climatology. Contour intervals are at 0.25 degrees Celsius.	62
Figure 2.10(ii)	Decadally averaged zonal mean temperatures from V2 of the edited HadRT 2.1s data, following quality control analysis, as anomalies from 1971-90 climatology. Contour intervals are at 0.25 degrees Celsius.	63
Figure 2.11	Figure showing decadally averaged temperatures at the 850hPa level in the original HadRT version. Regions edited in the final dataset are denoted by boxes.	64
Figure 3.1	Changes in areal grid box coverage of the HadRT dataset over the period 1958-1998. The key in the top right hand corner denotes the pressure levels. Note how, below an altitude of 50hPa, the total number of observations is generally consistent between the levels. Decreases in data availability occur both early in the record due to a lack of observations, and late in the record due to delays in release of the data.	87
Figure 3.2a	Global averages for the HadRT2.1s temperatures and HadCM2 ensemble mean fields. Note that expectations are that model fields will exhibit less variability.	88
Figure 3.2b	Global averages for the HadRT2.1s temperatures and HadCM3 ensemble mean fields. Note that expectations are that model fields will exhibit less variability.	89

Figure 3.3	Difference in zonally averaged temperatures between 1963-1972 and 1983-1992. The top two panels show HadRT2.1s and HadRT2.1 fields, the differences between these datasets being relatively small. Remaining panels indicate ensemble mean model responses to various anthropogenic forcing scenarios for HadCM2 (left hand panels) and HadCM3 (right hand panels).	90
Figure 3.4	Difference in temperatures at the 100hPa level between 1963-1972 and 1983-1992. The top two panels show HadRT2.1s and HadRT2.1 fields, the differences between these datasets being relatively small. Remaining panels indicate ensemble mean model responses to various anthropogenic forcing scenarios, and the respective model σ fields for HadCM2 (left hand panels) and HadCM3 (right hand panels). In these model fields those points outside 3σ from the observations are boxed.	91
Figure 3.5	Difference in temperatures at the 500hPa level between 1963-1972 and 1983-1992. The top two panels show HadRT2.1s and HadRT2.1 fields, the differences between these datasets being relatively small. Remaining panels indicate ensemble mean model responses to various anthropogenic forcing scenarios, and the respective model σ fields for HadCM2 (left hand panels) and HadCM3 (right hand panels). In these model fields those points outside 3σ from the observations are boxed.	92
Figure 3.6	Difference in an entire troposphere (300-850hPa) lapse rate diagnostic between 1963-1972 and 1983-1992. The top two panels show HadRT2.1s and HadRT2.1 fields. Remaining panels indicate ensemble mean model responses to various anthropogenic forcing scenarios, and the respective model σ fields for HadCM2 (left hand panels) and HadCM3 (right hand panels). In these model fields those points outside 3σ from the observations are boxed.	93
Figure 3.7	Best estimate and range of skill scores for a three-decade diagnostic. The best estimate is denoted by a square. There is no reason to expect the uncertainty range to be symmetric about this estimate. An ensemble is concluded to exhibit skill if its uncertainty range does not encompass zero. Confidence is increased with greater separation. In this approach a signal would be detectable if its best estimate was outside the uncertainty range of the null hypothesis no change field.	94

Figure 3.8	Best estimate and range of skill scores for a four-decade diagnostic. The best estimate is denoted by a square. There is no reason to expect the uncertainty range to be symmetric about this estimate. An ensemble is concluded to exhibit skill if its uncertainty range does not encompass zero. Confidence is increased with greater separation. In this approach a signal would be detectable if its best estimate was outside the uncertainty range of the null hypothesis no change field.	95
Figure 4.1	(after Hasselmann, 1976, and Mitchell and Karoly et al., 2001) Figure showing a simplified rationale for rotation of signals in climate change detection studies. The ellipse shows the 95% distribution of natural variability in the parameter in the simple two dimensional space. The black arrow indicates the original signal, the red arrow the rotated signal such that SNR is maximised. Although the red arrow (OC) has a lower absolute magnitude than the black arrow (OB), searching in this direction of phase space is more likely to yield a positive detection result as OC/OC_n is of greater magnitude than OB/OB_n . In the true climate system detection algorithms rotation is performed in n-dimensional space (where n is a large number).	132
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.	162
Figure 5.2	The effects of implementing the different weighting schemes on zonal-mean temperatures for V2 of the HadRT2.1s observations. The patterns have been normalised to enable a meaningful visual comparison to be made.	163
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.	164
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.	165
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.	166

Figure 6.1	Raw lower tropospheric HadRT2.1s V2 input temperature fields. The superimposed boxes indicate the choice of regions in the preferred 10-area “smart” LAA input diagnostic used in the current study.	229
Figure 6.2	10-area “smart” LAA input diagnostic for Lower Tropospheric HadRT2.1s V2 temperatures.	230
Figure 6.3	5-area “smart” LAA input diagnostic for Lower Tropospheric HadRT2.1s V2 temperatures.	231
Figure 6.4	Changing beta estimates with increasing truncation for HadCM2 upper tropospheric temperatures. The best-guess model signal amplitude estimate in the observations is given by the bold line, with 90% uncertainty ranges denoted by grey shading. Detection confidence limits are denoted by a red line. Where the residuals are inconsistent this is marked by an asterisk. Results are considered for three input signal combinations, and four choices of LAA diagnostic.	232
Figure 6.5	OLS regression ellipses for layer average temperature diagnostics for HadCM2 at truncation 21. The cross represents the best-guess amplitude estimator in each case, and the ellipse the 90% confidence interval as to the potential value of the true solution.	233
Figure 6.6	Changing beta estimates with increasing truncation for HadCM3 upper tropospheric temperatures. The best-guess model signal amplitude estimate in the observations is given by the bold line, with 90% uncertainty ranges denoted by grey shading. Detection confidence limits are denoted by a red line. Where the residuals are inconsistent this is marked by an asterisk. Results are considered for three input signal combinations, and four choices of LAA diagnostic.	234
Figure 6.7	Reconstructed global mean temperature trends of upper tropospheric temperatures for HadCM2 and HadCM3. The “observations” are projections onto leading modes of model simulated internal variability, and therefore differ between models. In each case the reconstruction is based upon the signals multiplied by their best-guess amplitude estimates.	235

Figure 6.8	Changing beta estimates with increasing truncation for HadCM2 lower tropospheric temperatures. The best-guess model signal amplitude estimate in the observations is given by the bold line, with 90% uncertainty ranges denoted by grey shading. Detection confidence limits are denoted by a red line. Where the residuals are inconsistent this is marked by an asterisk. Results are considered for three input signal combinations, and four choices of LAA diagnostic.	236
Figure 6.9	Changing beta estimates with increasing truncation for HadCM3 lower tropospheric temperatures. The best-guess model signal amplitude estimate in the observations is given by the bold line, with 90% uncertainty ranges denoted by grey shading. Detection confidence limits are denoted by a red line. Where the residuals are inconsistent this is marked by an asterisk. Results are considered for three input signal combinations, and four choices of LAA diagnostic.	237
Figure 6.10	Reconstructed global mean temperature trends of lower tropospheric temperatures for HadCM2 and HadCM3. The “observations” are projections onto leading modes of model simulated internal variability, and therefore differ between models. In each case the reconstruction is based upon the signals multiplied by their best-guess amplitude estimates	238
Figure 6.11	Changing beta estimates with increasing truncation for HadCM2 near-surface temperatures. The best-guess model signal amplitude estimate in the observations is given by the bold line, with 90% uncertainty ranges denoted by grey shading. Detection confidence limits are denoted by a red line. Where the residuals are inconsistent this is marked by an asterisk. Results are considered for three input signal combinations, and four choices of LAA diagnostic.	239
Figure 6.12	Changing beta estimates with increasing truncation for HadCM3 near-surface temperatures. The best-guess model signal amplitude estimate in the observations is given by the bold line, with 90% uncertainty ranges denoted by grey shading. Detection confidence limits are denoted by a red line. Where the residuals are inconsistent this is marked by an asterisk. Results are considered for three input signal combinations, and four choices of LAA diagnostic.	240

Figure 6.13	Reconstructed global mean temperature trends of near-surface temperatures for HadCM2 and HadCM3. The “observations” are projections onto leading modes of model simulated internal variability, and therefore differ between models. In each case the reconstruction is based upon the signals multiplied by their best-guess amplitude estimates.	241
Figure 6.14	Changing beta estimates with increasing truncation for HadCM2 free troposphere lapse rates. The best-guess model signal amplitude estimate in the observations is given by the bold line, with 90% uncertainty ranges denoted by grey shading. Detection confidence limits are denoted by a red line. Where the residuals are inconsistent this is marked by an asterisk. Results are considered for three input signal combinations, and four choices of LAA diagnostic.	242
Figure 6.15	Changing beta estimates with increasing truncation for HadCM3 free troposphere lapse rates. The best-guess model signal amplitude estimate in the observations is given by the bold line, with 90% uncertainty ranges denoted by grey shading. Detection confidence limits are denoted by a red line. Where the residuals are inconsistent this is marked by an asterisk. Results are considered for three input signal combinations, and four choices of LAA diagnostic.	243
Figure 6.16	Reconstructed global mean temperature trends of free troposphere lapse rates for HadCM2 and HadCM3. The “observations” are projections onto leading modes of model simulated internal variability, and therefore differ between models. In each case the reconstruction is based upon the signals multiplied by their best-guess amplitude estimates.	244
Figure 6.17	Changing beta estimates with increasing truncation for HadCM2 entire troposphere lapse rates. The best-guess model signal amplitude estimate in the observations is given by the bold line, with 90% uncertainty ranges denoted by grey shading. Detection confidence limits are denoted by a red line. Where the residuals are inconsistent this is marked by an asterisk. Results are considered for three input signal combinations, and four choices of LAA diagnostic.	245

Figure 6.18	Changing beta estimates with increasing truncation for HadCM3 entire troposphere lapse rates. The best-guess model signal amplitude estimate in the observations is given by the bold line, with 90% uncertainty ranges denoted by grey shading. Detection confidence limits are denoted by a red line. Where the residuals are inconsistent this is marked by an asterisk. Results are considered for three input signal combinations, and four choices of LAA diagnostic.	246
Figure 6.19	Reconstructed global mean temperature trends of entire troposphere lapse rates for HadCM2 and HadCM3. The “observations” are projections onto leading modes of model simulated internal variability, and therefore differ between models. In each case the reconstruction is based upon the signals multiplied by their best-guess amplitude estimates.	247
Figure 6.20	Changing beta estimates with increasing truncation for HadCM2 lower troposphere lapse rates. The best-guess model signal amplitude estimate in the observations is given by the bold line, with 90% uncertainty ranges denoted by grey shading. Detection confidence limits are denoted by a red line. Where the residuals are inconsistent this is marked by an asterisk. Results are considered for three input signal combinations, and four choices of LAA diagnostic.	248
Figure 6.21	Changing beta estimates with increasing truncation for HadCM3 lower troposphere lapse rates. The best-guess model signal amplitude estimate in the observations is given by the bold line, with 90% uncertainty ranges denoted by grey shading. Detection confidence limits are denoted by a red line. Where the residuals are inconsistent this is marked by an asterisk. Results are considered for three input signal combinations, and four choices of LAA diagnostic.	249
Figure 6.22	Reconstructed global mean temperature trends of lower troposphere lapse rates for HadCM2 and HadCM3. The “observations” are projections onto leading modes of model simulated internal variability, and therefore differ between models. In each case the reconstruction is based upon the signals multiplied by their best-guess amplitude estimates.	250

Figure 6.23	Changing beta estimates with increasing truncation for HadCM2 considering only the “smart” 10-area LAA diagnostic for all six input temperature variables. For each variable only the best-guess estimate is shown (see key for variables). The area of grey shading indicates regions where all six uncertainties overlap entirely. When the residuals are found to be inconsistent for any variable, the line is shown in feint.	251
Figure 6.24	Plot assessing the likely sensitivity of principal results for HadCM2 for the three layer average temperature variables to choice of regression algorithm. OLS results are shown as solid lines and TLS as dashed.	252
Figure 6.25	Raw observational input fields, optimised observational data, and best-guess reconstruction for near-surface temperatures. The reconstruction is based upon the G + S forcing combination for HadCM2.	253
Figure 6.26	Changing beta estimates with increasing truncation for HadCM3 considering only the “smart” 10-area LAA diagnostic for all six input temperature variables. For each variable only the best-guess estimate is shown (see key for variables). The area of grey shading indicates regions where all six uncertainties overlap entirely. When the residuals are found to be inconsistent for any variable, the line is shown in feint.	254
Figure 7.1	Illustration of the likely effects of relaxing from an ideal world to a real world situation, and results for HadCM2 G + S signal combination. The degree of separation in the hypothetical examples is a tuneable plot parameter and therefore highly uncertain. In each case only the 90% confidence interval is plotted. Key value abbreviations are given in chapter 6.	263

List of Tables

Table 1.1	Summary of HadCM2 and HadCM3 ensembles giving the acronyms and the run dates. A detailed description is given in Section 1.3 and Stott et al., 2001 (HadCM2) and Tett et al., 2001 (HadCM3).	31
Table 2.1	Location of potentially dubious grid box values, by season and annually, from near neighbour comparisons using version 2 of HadRT2.1s. Those locations exhibiting solely individual Z-Score errors are denoted by X, all others are denoted by X .	50-51
Table 2.2	Table defining the temporal sampling characteristics of flagged and unflagged grid boxes resulting from near-neighbour comparisons on both an annual and a seasonal basis. The tests used here are the student's t-test and standard F-test. Values of significance are quoted only when 90% or greater.	51
Table 2.3	Listing of grid boxes found to be anomalous by near neighbour comparisons, detailing potential reasons from available metadata. For more details on the status and potential deficiencies in the metadata see Gaffen (1996).	52-53
Table 2.4	Listing of grid boxes found to be anomalous in data sparse regions, detailing potential reasons from available metadata. For more details on the status and potential deficiencies in the metadata see Gaffen (1996).	53
Table 3.1	Percentage of individual grid box values greater than 3σ from the observations for each temperature variable. Values highlighted in yellow indicate model skill compared to the null hypothesis field of "no change".	84
Table 3.2	RMSD values between modelled and observed fields for each temperature variable. Values highlighted in yellow indicate model skill compared to the null hypothesis of "no change".	85
Table 3.3	Overall skill scores for the model fields when compared to a null hypothesis of zero change. For any level or lapse rate to gain a score it must exhibit skill both at the grid-box level and in the RMSD statistic. The maximum possible score is, therefore, 12. The fields are split so as to consider skill scores and their uncertainty ranges for temperatures on pressure levels and lapse rate diagnostics separately. Model fields considering anthropogenic influences consistently provide the best explanation, although there are also likely solar, and	86

possibly volcanic influences, and “no change” cannot be ruled out. The choice of decadal averaged data is probably sub-optimal when searching for natural external forcing signals.

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).	153
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.	154
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.	155
Table 5.4	As Table 5.3, except considering the results from a Total Least Squares regression methodology.	156
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.	157
Table 5.6	As Table 5.2, except for a troposphere-only input diagnostic.	158
Table 5.7	As Table 5.3, except for a troposphere-only input diagnostic.	159

Table 5.8	As Table 5.4 for a troposphere-only input diagnostic.	160
Table 6.1	Summary of the individual candidate forcings considered in this study and their acronyms (HadCM3 forcings are aliased to the HadCM2 forcings). For more information on how these forcings are prescribed within the models see chapter 1 (section 1.3), Mitchell et al. (1995), and Tett et al. (2001).	218
Table 6.2	Detection results for all possible 1-, 2-, and 3- input signal combinations for upper tropospheric temperatures. . If a signal is detected in any particular input combination then it is recorded in this table, along with whether it is significantly over-(\$) or under-(*) estimated in its amplitude (inconsistent) within the model. In those cases where signals are detected but residuals are inconsistent before the maximum truncation of 21 they are highlighted in orange. Other cases of early residual consistency test failure are not indicated here.	219
Table 6.3	As Table 6.2 except considering lower tropospheric temperatures.	220
Table 6.4	As Table 6.2 except considering near-surface temperature series.	221
Table 6.5	As Table 6.2 except considering free troposphere (UT-LT) lapse rates.	222
Table 6.6	As Table 6.2 except considering entire troposphere (UT-Surf) lapse rates.	223
Table 6.7	As Table 6.2 except considering lower tropospheric (LT-Surf) lapse rates.	224
Table 6.8	SNR values at truncation 11 for the individual signals being considered. Those cases where the signal is indistinguishable from noise at the 90% confidence interval are highlighted in orange.	225
Table 6.9	SNR values at truncation 21 for the individual signals being considered. Those cases where the signal is indistinguishable from noise at the 90% confidence interval are highlighted in orange.	226
Table 6.10	Summary of principal results for HadCM2. The scoring	227

system is subjective, but is based upon objective OLS detection analyses, and therefore the overall pattern shown is likely to be robust to other choices of scoring mechanism over a reasonable range. Table considers signal detection and consistency (attribution), regression residuals consistency, and how well the temporal trends for 1960-1995 are reproduced at the global scale.

Table 6.11 As Table 6.10, except considering results for HadCM3. 228