4. Optimal detection methodologies

In this chapter, current optimal multi-variate statistical approaches to detection and attribution are introduced. No attempt is made to develop further any previously published detection methodologies. Rather, the aim is to provide the information necessary to understand the analysis in subsequent chapters. Univariate (Stage I) approaches are not discussed as they have extremely limited scope in yielding either unambiguous detection or attribution (Santer et al., 1996, chapter 1). Nor are the principles behind pattern correlation techniques (Santer et al., 1993, 1995), although the linkage between these and optimal regression is discussed. The principal limitations of the optimal detection methodologies currently employed are also discussed. Section 4.1 introduces the components necessary for an optimal detection exercise, along with some of the assumptions common to all approaches. In Section 4.2 the optimal regression approaches of Allen and Tett (1999) (henceforth AT99) and Allen and Stott (2001) (henceforth AS01), proposed independently by Levine and Berliner (1999), are outlined, and their limitations and potential applications discussed. These methodologies are concentrated upon as they are the detection algorithms implemented in chapters 5, 6, and 7. Section 4.3 briefly describes other multi-variate detection approaches pursued to date, couching them where possible in terms of the classical regression approaches detailed in section 4.2. The principal results of, and caveats that apply to, all of these methodologies have been discussed previously in chapter 1, hence solely the technical details of detection are focused upon here. Section 4.4 concludes.

4.1 The basics of climate change detection

Climate change detection can only ever be addressed as a statistical problem, whereby probabilistic statements are made as to the most likely causes of recently observed climate change. It will never be possible to prove absolutely all the causes of recently observed climate change as the Earth's climate system is highly complex.

Not only is the system highly complex, it is also poorly observed and crudely modelled (see chapter 1). There are numerous forcings that can affect the evolution of the climate system, only a few of which (albeit the most likely major candidate forcings) have been considered to date in climate models. Furthermore, the atmosphere is coupled to the oceans, the cryosphere and the biosphere. Small perturbations to any component of this coupled system will cause the climate system to vary, sometimes in preferred modes (NAO, ENSO, QBO for example) – these variations are termed 'natural internal variability' or 'noise' in detection studies, and exist on all space and time scales. The detection of the physical manifestation of any large-scale forcing (such as anthropogenic CO_2 emissions due to the combustion of fossil fuels) upon the global climate system is, therefore, some form of statistical "emerging signal in noise" problem. To quantify the likelihood of any forcing signal(s) (e.g CO_2) influencing the global climate system, three components are required in any detection study:

- An observational realisation of the spatio-temporal evolution of a climate parameter such as near-surface temperature.
- A modelled realisation of the spatio-temporal response of that climate parameter to the candidate forcing signal(s).
- An estimate of the spatio-temporal natural variability structure of that climate parameter (with no external forcings upon the climate system), from which to claim significance of any statistic.

Those modelled and observed datasets used in the current thesis are discussed in detail in chapter 1. In all detection approaches to date, the observations are compared to the model-derived signals to yield a statistical indicator of their similarity. The model-derived signals are also compared to model estimates of natural internal climate variability. These can be derived from either a section(s) of model control run (run with no external forcings), or an estimate based upon intra-ensemble variability. This allows for an estimate as to the natural variability of the statistic, from which to claim whether or not the observed result is significantly different to

that expected by chance alone. Therefore, any claim of signal detection is **critically** dependent upon the adequacy of the model control estimate of natural internal variability. Assumptions must also be made regarding the veracity of both the observations and the model-derived signals. Most of these assumptions can be directly assessed and quantified through the application of sensitivity studies (Tett et al., 1999, Gillett et al., 2000a, for example). Additionally, there are many other studies which have considered the realism of both observed and modelled variables and the spatial and temporal scales at which both the models and the observations are adequate (Stott and Tett, 1998, Gillett et al., 2000, Johns et al., 1997, NRC, 2000, for example).

Most recent detection studies have employed an "optimal" detection methodology in an attempt to maximise the chances of a successful outcome. A number of different approaches have been proposed in these optimal detection studies. The rationale behind all such approaches is to rotate the searched-for model-derived signal(s) and the observations in such a way as to maximise the SNR. Those regions of phase space where the signal(s) is (are) dominant over climate noise due to natural variability (estimated from a segment of model control) are given higher weighting in the detection algorithm. Conversely, those regions of phase space where the noise is dominant over the signal(s) are given lower weighting. This is seen most easily when considering a very simple hypothetical system that can only vary in two dimensions (the true climate system varies in an infinite dimensional phase space). Figure 4.1 illustrates graphically how rotation increases the SNR in this hypothetical system (after Hasselmann, 1976, see Mitchell and Karoly et al., 2001). Rotation of the signal is invariably carried out with respect to a realisation of natural variability from a model control run as suitable observational datasets do not exist, either being potentially polluted by the searched-for signals, or containing residual uncertainties (Jones and Hegerl, 1998). There is an associated increase in uncertainty involved in the optimal detection statistic if the estimate of natural variability in the climate parameter is (grossly) incorrectly diagnosed within the model control. Importantly, it has been shown that this can only ever lead to a conservative statistical error under

an optimal detection approach (AT99, see section 4.2.1, Hasselman, 1997, Hegerl et al., 1997).

AT99, AS01, and Levine and Berliner (1999) have shown that the optimal detection approach of Hasselmann (1997) can be couched in terms of multi-linear regression. This is the approach employed in this thesis and hence concentrated upon in this chapter. Hegerl and North (1997) have further shown how a number of independent approaches that have been taken to optimal detection, including that of Hasselmann (1997), are broadly comparable (see section 4.3 for more details of these methodologies, and how they relate to the regression approach outlined in section 4.2). A regression approach has advantages over these other optimal detection approaches in terms of being able to explicitly quantify model signal amplitudes in the observations. The residuals of the regression can also be compared to independent realisations of natural variability as a consistency test. By quantifying signal strength in a meaningful manner, it potentially becomes possible to constrain future climate projections under a transient forcing (Allen et al., 2000, 2001). A multiple linear regression approach is also the Uniformly Most Powerful (UMP) detection statistic (Levine and Berliner, 1999) in a classical frequentist statistical approach, although more powerful Bayesian alternatives may exist. Optimal multi-linear regression approaches can be shown to be a special case Bayesian approach (Allen et al., 2001), whereby the prior probability density function is uniform (although the proof is not detailed here, it is useful to note the equivalence). Incorporating a Bayesian statistical approach might increase the power of future detection algorithms (Berliner et al., 2000). Two variants of the multiple linear regression technique have emerged, Ordinary Least Squares (henceforth referred to as OLS) (AT99, Levine and Berliner, 1999, see section 4.2.1), and Total Least Squares (henceforth referred to as TLS) (AS01, see section 4.2.2). Most published studies to date using a multi-linear regression approach have considered results from OLS regression.

There are numerous potential viewpoints on how exactly the detection and attribution problem should be treated in a statistical framework (Levine and Berliner, 1999). It should be noted that if the climate system is treated in the wrong framework then potentially erroneous results could ensue unless adequate "safety checks" are performed upon the results. Levine and Berliner (1999) characterise the approach of regression as taking an ensemble-stochastic view, rather than an intrinsic-stochastic view, which they argue is the truly valid approach. In an intrinsic-stochastic view the observations are assumed to be made up of a number of components:

- The true signal.
- Uncertainties due to ensemble effects. A small, finite number, of model ensemble members are started under imperfectly known initial conditions, adding uncertainty to the estimated signal. The signal estimate will consist of both the true signal and an additional residual noise term.
- Operational uncertainties whereby all estimates are ascribed probability distributions, since the model is not a perfect representation of the real world, and the observations are recognised as only a representation of the true variable values.
- Unexplained noise.

The regression algorithms proposed by AT99 and AS01 (see section 4.2) explicitly account for ensemble effects only (hence ensemble-stochastic), but propose a consistency check on the residuals, which would identify, although not necessarily ascribe causes to, gross errors arising from the latter two sources. How one views the system is, therefore, in practice unlikely to greatly affect the results of those optimal regression studies that contain a consistency check on the residuals.

Optimal detection algorithms to date have only generally considered changes in the first order statistics of variables (Levine and Berliner, 1999). That is, these studies have made distributional assumptions about the searched-for atmospheric parameters. The approach is to assume that natural variability has a fixed

distribution, and that a transient forcing elicits a change solely in the mean of the distribution (shifting the entire distribution by a value equal to the change in the mean), without impacting upon the higher order statistics. However, it should be noted that there is no physical or theoretical reason why higher order statistics of variables should remain stationary under a transient climate change. Timmerman (1999) outlines a potential optimal detection approach using variance and skewness (second and third order statistics respectively) characteristics, and applies this to a study considering potential non-stationary ENSO characteristics. There is no reason, at least in theory, as to why the optimal regression approach outlined in section 4.2 could not be advanced to consider such statistics. However, it is highly unlikely that HadRT upper air temperatures (Parker et al., 1997) are a suitable candidate variable for such a study, both due to the relatively short observational record available, and remaining observational dataset uncertainties (chapter 2). Furthermore, the sampling error, for a given sample size, will be larger for higher order statistics, so the small sample HadRT dataset would seem to be inadequate. Advancement of such an approach is therefore left to others.

4.2 Detection seen as optimal regression

4.2.1 Ordinary Least Squares Regression

4.2.1.1 The basic regression model

AT99 and AS01 apply the general statistical definition of multi-linear regression as follows. The basic equation to be solved is:

$$y = X\beta + \nu \qquad (1)$$

Where: y is a realisation of the observations (a column vector of rank *l*).

X is a set of *n* response patterns to plausible forcing mechanisms, each of rank

l, denoted as the columns of the ($n \times l$) array.

β is a row vector containing the amplitude estimators (n multipliers) of the columns of X in y.

ν is a rank l column vector of the climate noise (residuals)

It is not necessary for y to be the raw observations, so long as X, y and ν are treated identically within the procedure. For example, in previous studies using an OLS regression algorithm, looking at near-surface temperatures (Tett et al., 1999, 2001, Stott et al., 2001), a space-time approach has been undertaken, whereby the input vectors are treated from a 3-D discrete latitude, longitude, time array. Decadal averages are taken and converted into a spherical harmonic coefficient (SHC) representation truncating at wave-number 5 such that each decade consists of 25 values (Stott and Tett, 1998 provide a justification for this cut-off value). This SHC representation of each decadal average is then used as input to the OLS regression, stacked in time, to yield a space-time input diagnostic vector. The reader is directed to Tett et al. (1999, 2001), and Stott et al. (2001) for further details of this pre-processing. In principle, there is no reason why alternative forms of pre-processing cannot also be employed to yield the input vectors. Those pre-processing algorithms employed in this thesis are described in chapters 5 and 6.

Under an OLS regression approach, the climate noise covariance, which is required to undertake the optimisation, is given by:

$$C_N \equiv \varepsilon\left(\nu\nu^T\right) \tag{2}$$

where ε is the expectation operator. C_N is generally unknown, as the climate system has not been observed in sufficient detail on the multi-centennial to millennial timescales that are required to derive an accurate estimate of multi-decadal natural internal climatic variability important in detection studies. Any such observationally based estimate would also likely contain externally forced variability due to solar variations, volcanic eruptions, and more recent anthropogenic influences, amongst others. Although these effects could be estimated (e.g. by climate modelling), and removed (Jones and Hegerl, 1998), it is highly unlikely that the corrections will be

perfectly known, especially previous to the most recent 30 years. Therefore, the variability in the resulting dataset will not solely be due to natural variability, and C_N must invariably be estimated from a segment of model control run, the assumption being that the model is a reliable estimator of natural internal climate variability.

In the true climate system natural internal variations are not WHITE (independent, identically distributed) noise in either space or time. The system exhibits distinct correlations (covariance), whereby it tends to have preferred modes over a large range of space and timescales. If this were not taken into account, the resulting estimates of β from (1) would be both highly inefficient (as the rotation would be sub-optimal), and have strongly biased error estimates (AS01). The regression can be optimised by noting that the best linear unbiased (BLUE) estimator of β can be found by introducing a pre-whitening operator, P, whereby the transformed noise estimate, Pν, appears to be WHITE in nature such that:

$$\varepsilon\left(P\nu\nu^T P^T\right) = PC_N P^T = I \qquad (3)$$

where I is the identity matrix. The BLUE estimator for β varies less between experimental realisations than any other linear unbiased estimator. It is also the solution that minimises the variance in the solution. That is, if one were actually able to repeat the entire experiment with new observations and model predictions, the BLUE estimator would vary less between these repeat experiments than any other unbiased linear estimator (AT99). This pre-whitening procedure could also be considered in terms of SNR, or an explicit Bayesian treatment. Differences, however, are effectively a matter of interpretation, rather than fundamental approach (AT99). These different approaches to optimisation may yield very different tools for subsequent consideration of the results. The OLS regression approach advocated by AT99 has advantages over those optimal techniques used previously (see Hegerl and North, 1997, and section 4.3), as it assumes that the pattern is correct (with an uncertain amplitude). Previous approaches have been unable to distinguish between the influence of pattern and amplitude upon detection results (Levine and Berliner, 1999). In the presence of a finite ensemble size, there will be uncertainty in the pattern as well as the amplitude of the signal responses (Barnett et al., 2000). Hence

the OLS regression approach is only an approximation, especially in the case of small population ensembles and weak signals (AS01). Furthermore, it makes the tacit assumption that the climate system is linear on the time and space scales considered, such that any model derived climate response to a forcing is able to be directly linearly scaled to fit the observations.

The BLUE estimator is satisfied if $P^T P = C_N^{-1}$, provided such an inverse exists (i.e. C_N is not under-sampled and therefore non-invertible), and the Gauss-Markov theorem (Mardia et al., 1979) may be invoked as Pν is indistinguishable from white noise. To prove that the best (lowest variance) linear unbiased estimator for β is BLUE, the following merit function must be minimised (AS01):

$$r^2\left(\tilde{\beta}\right) = \left(PX\tilde{\beta} - Py\right)^T \left(PX\tilde{\beta} - Py\right) = \tilde{\nu}^T P^T P \tilde{\nu} \quad (4)$$

where ~ indicates the best-guess values, and (4) aims to minimise the residuals in the observations (geometrically viewed as minimising the squared differences in the y direction (dependent variable in classical terms) – hence OLS). The minimum condition, r^2_{min}, is satisfied assuming:

$$\tilde{\beta} = \left(X^T P^T P X\right)^{-1} X^T P^T P y = \left(X^T C_N^{-1} X\right)^{-1} X^T C_N^{-1} y \equiv F^T y \quad (5)$$

Where:

$$F^T = \left(X^T C_N^{-1} X\right)^{-1} X^T C_N^{-1} \quad (6)$$

F^T is the term that extracts the BLUE signal estimator from the observations (y). This is the OLS regression solution applied to the transformed (pre-whitened) variables PX and Py (AT99). F^T is called the BLUE estimator in classical regression literature (Mardia et al., 1979), but for consistency with previous detection studies, AT99 suggest that the *m* rows of F^T should be termed the "distinguishing fingerprints".

As the estimate of the signal amplitudes is unbiased $\varepsilon\left(\tilde{\beta}\right) = \beta$, and the $m \times m$ covariance matrix $\varepsilon\left[\left(\tilde{\beta} - \beta\right)\left(\tilde{\beta} - \beta\right)^T\right]$ is given by the inverse of the Hessian matrix:

$$V\left(\tilde{\beta}\right) \equiv \varepsilon\left(\left(\tilde{\beta}-\beta\right)\left(\tilde{\beta}-\beta\right)^{T}\right)$$
$$V\left(\tilde{\beta}\right) = \left[\frac{1}{2}\frac{\partial^{2}\left(r^{2}\right)}{\partial\left(\beta'\right)^{2}}\right]^{-1} \quad (7)$$
$$V\left(\tilde{\beta}\right) = \left(X^{T}C_{N}X\right)^{-1}$$

where V is the variance (AS01). Therefore, ignoring any uncertainty in the natural variability estimate, C_N, the variance can be rewritten as:

$$V\left(\tilde{\beta}\right) = F^{T}C_{N}F \quad (8)$$

It should be noted that the uncertainty estimate in the solution is independent of the observations (y) in undertaking an OLS regression approach. **If ν is multivariate normal** (which it is following the application of the pre-whitening operator, P) then this implies that:

$$\left(\tilde{\beta}-\beta\right)^{T}\left[V\left(\tilde{\beta}\right)\right]^{-1}\left(\tilde{\beta}-\beta\right) = \Delta r^{2}\left(\beta\right) \equiv r^{2}\left(\beta\right) - r_{\min}^{2} \approx \chi_{m}^{2} \quad (9)$$

The regression solution can be visualised as a cloud of equally plausible solution points each comprising of the *m* values of β (signal amplitudes) in *m*-dimensional signal phase space co-ordinates. Confidence limits on the distribution can be found by calculating critical χ_m^2 values, and evaluating the corresponding values of β for the LHS in equation (9) for the *m* distinguishing fingerprints being considered (AT99). The true value of β is then expected to lie within the *m*-dimensional limits defined at the prescribed confidence interval. A number of confidence intervals can be considered simultaneously to yield isopleths of a PDF (probability density function) that the true solution lies within each bounding isopleth of the PDF. The PDF illustrates the uncertainty in the estimates of the scaling required on the m distinguishing fingerprints, F^T, to fit the observations, y. A subset of signals can be considered by extracting the relevant columns from X, and evaluating the LHS of equation (9) with the reduced number of degrees of freedom (m) for χ_m^2.

Any estimate of the variance of $\tilde{\beta}$ gained from equation (8) can also be used to yield potentially useful information on model characteristics. AT99 note that knowledge of the variance of $\tilde{\beta}$ can be used to estimate the implied uncertainty in any scalar linear

diagnostic of the climate system, ϕ. This diagnostic can, with trivial exceptions, be represented as a projection of the observations onto a vector of weights, w, such that $\phi = w^T y$. What specifically is being considered will depend upon the elements of w. Neglecting any uncertainty in X, the variance of ϕ attributable to uncertainty in the amplitude estimate $\tilde{\beta}$ is given by:

$$V(\phi) = w^T X V(\tilde{\beta}) X^T w \quad (10)$$

The potential therefore exists to consider the best-guess values and associated uncertainties for different solutions to ϕ, where ϕ represents for example global means (Allen et al., 2000, 2001), or regional averages. Regression results can then be used to advance from a relatively simple detection exercise towards considering which model aspects are consistent with the observations and, therefore, to identify likely model errors (AT99). Care must be taken to ensure against introducing a common factor effect (Santer et al., 1993), whereby results from such a study are subsequently fed back into the model, making it more like the observations for no physically plausible reason.

The OLS regression methodology detailed above is true in the limit of an infinite ensemble, when the estimate of X will be noise free. However, in real world detection problems, a very small finite ensemble size is available, and hence X will tend to bias signal amplitude estimates $(\tilde{\beta})$ towards zero (Mardia et al., 1979, Bell, 1986), and increase the true variance in the estimator by a factor of $\approx$ 1+1/m, where m is the number of ensemble members. AT99 show how this variance factor correction is applied to OLS regression approaches assuming that the shapes of the columns in X are approximately constant, whilst their amplitude changes due to sample size effects. It should be noted that this is solely an approximation, as expectations would also be for changes in the signal shape due to finite sampling (Barnett et al., 2000), but these changes are considered likely to be much smaller, at least for strong signals (AT99). Stott et al. (2001) discuss how this correction factor is only necessary when considering signal consistency and PDF isopleth limits, and not individual signal detection.

4.2.1.2 Accounting for uncertainty in the model control variability estimate

The key problem in any optimal detection algorithm is that the noise covariance array C_N is unknown, and must be estimated from a model control run segment, which almost certainly does not capture all significant modes and scales of natural variability (Stott and Tett, 1998). Alternatively, the estimate could arise from model intra-ensemble variability which is assumed to arise solely from natural internal variability (Tett et al., 2001). If intra-ensemble variability is used, then the procedure is to remove the ensemble average value and scale each value by factor $\sqrt{(m-1)/m}$, where m is the number of ensemble members (Tett et al., 2001). The noise estimate is gained by evaluating:

$$\hat{C}_N = \frac{1}{n} Y_N Y_N^T \tag{11}$$

where the columns of Y_N represent successive independent realisations of pseudo-observations from the control run or intra-ensemble variability. These should mimic the available observations as closely as possible. In particular, a similar missing data mask should be applied (AT99). Typical trend lengths considered in detection studies are 30-100 years. Control run data generally exists for a maximum of 1000-2000 years for any model. Therefore, the columns of Y_N will generally contain many more spatial elements than there are independent realisations of natural variability, even in the spatio-temporally reduced input fields (for example SHC decadal average space (Tett et al., 1999, 2001)) being considered. Hence $\hat{C}_N$ is non-invertible.

Although $\hat{C}_N$ is non-invertible, this is not a problem as C_N^{-1} is not actually required for the estimate $\tilde{\beta}$ to be BLUE (AT99). The requirement is that P, the pre-whitening operator is satisfied, hence I, the unit matrix on the RHS of (3) need not be $l \times l$. If it is assumed that the control run adequately samples variability in the sub-space sampled

by the κ highest variance EOFs of $\hat{C}_N$, then one can consider a truncated pre-whitening operator $P^{(\kappa)}$. In such a construct, the rows of $P^{(\kappa)}$ are the κ highest variance EOFs of the control, weighted by their inverse singular values (so as to give weight to those regions of phase space with lower natural variability). Hence $P^{(\kappa)}\hat{C}_N P^{(\kappa)T}$ is equivalent to the rank κ identity matrix. Such an approach is equivalent to using the Moore-Penrose pseudo-inverse $P^{(\kappa)T}P^{(\kappa)}$ in place of C_N^{-1} (AT99).

Using this method has obvious potential implications on the probable outcome, successful or otherwise, of any detection exercise. The uncertainty estimate $V\left(\tilde{\beta}\right)$ is critically dependent upon the choice of κ. It will decrease in amplitude with increasing values for κ by construction, as those EOFs with smaller singular values (explaining less of the total variance) are incorporated. However, the available model control is not generally of sufficient length to adequately sample all modes of model estimated natural variability on the multi-decadal timescales considered in climate change detection. Therefore, it could be that poorly sampled EOFs, with artificially low estimates of natural variability (singular values) are incorporated. This could easily lead to a Type I statistical error (rejection of a valid null), if the choice of κ leads to the inclusion of modes under-sampled in the model control run, rather than modes with truly lower variance. These EOFs would be given unrealistically high weighting under the optimisation procedure.

Although $P^{(\kappa)}\hat{C}_N P^{(\kappa)T}$ is equivalent to the unit matrix, $\hat{C}_N \neq \mathrm{C_N}$ due to the finite length of the control segment under consideration, hence equation (3) is only partially satisfied using this approach (AT99). Equally, the estimate $P^{(\kappa)}$ is biased towards the section of control from which it is derived. An independent sample of control with estimated covariance $\hat{C}_{N_2}$ (derived in the same manner as equation (11)), will yield diagonal elements of $P^{(\kappa)}\hat{C}_{N_2} P^{(\kappa)T}$ which are generally <1 (AT99). This

introduces a bias into the covariance of any estimate of $\widetilde{\beta}$, such that the confidence limits of the distribution will be artificially small, increasing the chances of a Type I statistical error arising. These effects can be taken into account by employing separate sections of model control run for the optimisation and hypothesis testing components of detection (Hegerl et al., 1996, AT99). AT99 show how this is achieved in the regression framework by replacing equation (8) with the estimate:

$$\begin{aligned}\widetilde{V}\left(\widetilde{\beta}\right) &= \hat{V}\left(F_1^T y_{N_2}\right) \\ &= \frac{1}{n_2} F_1^T Y_{N_2} Y_{N_2}^T F_1 \\ &= F_1^T \hat{C}_{N_2} F_1 \end{aligned} \tag{12}$$

As $\varepsilon\left(F_1^T y_{N_2}\right) = 0$, equation (12) is the standard estimate of the variance of $\widetilde{\beta}$ obtained by summing squares over n_2 realisations of $F_1^T y_{N_2}$. In the limit of an infinite control run, equation (12) will collapse to equation (8), as both covariance estimates tend towards the true covariance of the (model) system (AT99).

Considering uncertainty limits due to errors in the estimate for $\widetilde{\beta}$, equation (9) is replaced by:

$$\left(\widetilde{\beta} - \beta\right)^T \left[\widetilde{V}\left(\widetilde{\beta}\right)\right]^{-1} \left(\widetilde{\beta} - \beta\right) \equiv \varepsilon^2(\beta) \approx mF_{m,v} \tag{13}$$

a standard F distribution with m and v degrees of freedom in the numerator and denominator, with v being the number of degrees of freedom of $\hat{C}_{N_2}$, and m the number of signals under consideration. Similarly to equation (9), an *m*-dimensional PDF isopleth can be plotted by evaluating the loci of points relating to a given significance level of the F distribution (AT99). An F distribution will yield a larger uncertainty range than the equivalent Chi-squared distribution, although given a sufficiently large number of degrees of freedom (v), the F distribution will collapse toward the Chi-squared distribution, such that $\chi_m^2 = mF_{m,\infty}$. AT99 show that the limit is likely to be about v=100. There is no rigorous way of ascertaining the true degrees of freedom of the independent control section used in hypothesis testing. An

estimate of 1.5 times the number of non-overlapping chunks is advocated following analysis by Allen and Smith (1997). It should be stressed that this is solely an empirically-based estimate and may not hold true in all situations.

4.2.1.3 A consistency check on the residuals

An OLS regression approach yields a useful consistency test to ensure that the results are physically plausible, as the residuals (ν) can be checked against independent estimates of natural variability. To date such tests have assumed multivariate normality, which should be valid following the application of the pre-whitening operator, P. A null hypothesis is constructed that the model adequately represents the variability in the real world. This null explicitly must account for observational error as well as other potential sources of error outlined by Levine and Berliner (1999), despite these not being implicitly included in the regression model. The null is tested in the truncated phase space defined by the first κ EOF's in which the regression has been performed. AT99 propose a simple test of the null hypothesis whereby the residuals of the regression:

$$\tilde{\nu} = y - X\tilde{\beta} \tag{14}$$

would behave as mutually independent, normally distributed random (WHITE) noise in the coordinate system defined by $\hat{C}_N^{-1}$, such that:

$$r^2 = \tilde{\nu}^T \hat{C}_N^{-1} \tilde{\nu} \approx \chi^2_{\kappa-m} \tag{15}$$

is distributed as the sum of squares of κ-m normally-distributed random variables with a mean of 1. If the truncation, κ, increases such as to include model states with unrealistically low variance estimates in the control segment used for optimisation, then it will tend to distort uncertainty analysis in the regression. At the same time, the component orthogonal to X will tend to increase the residuals, such that the value of r^2 increases unrealistically. A critical value of the $\chi^2_{\kappa-m}$ distribution can be chosen as a threshold for rejection of the null hypothesis, providing a warning that the results of the regression are likely to be suspect. The assumption is that the control variability is adequate, and hence the test provides warning that uncertainty estimates are

becoming unreliable (see AT99 for more details). The test should, strictly speaking, be two-tailed as it can be argued that too good a fit to the observations (such that the residuals are an underestimate of natural variability) should carry the same cost as too poor a fit. In practice the test has been applied both as a one-tailed test (Allen et al., 2000, 2001, AT99), looking for values which exceed only the upper threshold, and a two-tailed test (Tett et al., 1999, 2001, Stott et al., 2001). There is no systematic bias in the results between the approaches. In the present analysis the one-tailed approach is used, as it is not desired to penalise too good a fit to the same extent as too poor a fit of the regression model to the observations.

There are potential limitations to the consistency test approach outlined above, especially when considering large values for κ. Remembering that the denominator in equation (15) $\left(\hat{C}_N\right)$ has been pre-whitened such that each element is unity, the hypothesis is that the numerator has a normal distribution about unity and, therefore, that the value of the solution to (15) will oscillate randomly around unity. If the denominator (truncation) becomes sufficiently large then introducing values systematically >1 into the numerator (based upon unrealistic control estimates), need not immediately trigger a failure of the test statistic. Furthermore, introducing a single large value >>1 randomly into the numerator, will be more likely to cause a test failure earlier in the truncation series, after which the statistic may become consistent again as the denominator becomes progressively proportionally larger. Although there remains scope for a more optimal approach to testing for consistency of the residuals, the current test is felt to be adequate for the purposes of this thesis, although its limitations should be borne in mind.

AT99 note that the consistency check should reduce to an F-test if separate sections of control are used in the optimisation and testing procedures.

$$\tilde{\nu}^T \hat{C}_{N_2}^{-1} \tilde{\nu} \approx (\kappa - m) F_{\kappa-m,\nu} \tag{16}$$

Results using this approach are likely to be dependent upon the estimated value of v, the "true" degrees of freedom (see earlier discussion). Regardless of the consistency

test employed, a truncation limit criteria for optimisation is advocated by AT99, whereby the truncation is limited to the maximum value for κ which does not fail this independent consistency check. Tett et al. (2001) state that the correct test to use to remain overall conservative in approach is an F-test at the 90% confidence interval cut-off point. Furthermore, Tett et al. (2001) show how the residuals must be scaled to account for finite ensemble size, scaling the left hand side of (15) by a factor of 1/(1+s), whilst assuming that it maintains the same distribution, where s is defined as $s = \sum_{i=1}^{n} \left(\tilde{\beta}_i / m_i \right)^2$, and m_i is the number of members in the ith ensemble.

4.2.1.4 Recombining input fields to yield individual signal amplitude estimates

In certain cases, input signals must be recombined, following input to the regression algorithm, to yield the individual signal amplitude estimates. This recombination is undertaken under the implicit assumption that the combined signal response is a linear combination of the individual signal responses (AT99, Tett et al., 2001). This assumption is difficult to assess in any robust way, as orders of magnitude more ensemble members would be required to prove the existence of non-linear interactions between the combined and individual forcing responses (AS01). It is a necessary assumption when the ensembles do not contain the pure signal, but rather the signals in some combination, as is the case for the anthropogenic ensembles in both HadCM2 and HadCM3. Taking HadCM2 fields as an example, there exist three anthropogenic ensembles:

- G (Well-mixed greenhouse-gases)
- GS (Well-mixed greenhouse-gases plus sulphate aerosols)
- GSO (Well-mixed greenhouse-gases, sulphate aerosols, and stratospheric ozone depletion)

Suppose that it were desired to gain an estimate of the amplitude of the model signal response to G and S separately. It would be necessary to undertake a consideration of both G and GS ensemble responses:

$$y = \widetilde{\beta}_G X_G + \widetilde{\beta}_{GS} X_{GS} + \nu \tag{17}$$

However, $\widetilde{\beta}_{GHG}$, the estimated strength of the G signal will be dependent on both components of the regression. The G signal strength estimate in (17) is the additional greenhouse response required to explain the observations above and beyond that component in the GS signal (Tett et al., 2001):

$$y = \widetilde{\beta}_G X_G + \widetilde{\beta}_{GS}\left(X_G + X_S\right) + \nu$$

$$y = \left(\widetilde{\beta}_G + \widetilde{\beta}_{GS}\right)X_G + \beta_{GS} X_S + \nu \tag{18}$$

Hence the solution is:

$$\widetilde{\beta}_{GHG} = \widetilde{\beta}_G + \widetilde{\beta}_{GS} \tag{19}$$

$$\widetilde{\beta}_S = \widetilde{\beta}_{GS} \tag{20}$$

The variance in $\widetilde{\beta}_{GHG}$ is simply the sum of the variances in the two components on the RHS of (19).

4.2.1.5 Addressing possible input signal degeneracy

Tett et al. (1999) flag a potential problem when considering results of a multivariate linear regression algorithm such as OLS. If a signal (or combination of signals) resembles a further signal (or combination of signals), then the input signals are degenerate, and results from the regression are likely to be compromised. Consideration of both signals (or combinations) will yield very different amplitudes to those when considering solely one of the signals (or combinations) in the regression algorithm. When considering both the degenerate signals (or combinations), a potentially large range of signal-amplitude combinations would yield a plausible explanation of the observations. This is likely to be a problem in the climate change detection framework, in which any distinct climate forcings are likely to elicit dynamic responses that will tend to reduce the differences between the

forcing and response patterns. Tett et al. (1999) apply three simple tests for degeneracy on the signal correlation matrix ($C=X^{*T}X^{*}$), where X^{*} is the matrix of optimal normalised signals considered in the regression, the columns of F^{T}. The tests used are from Mardia et al. (1979):

1. Threshold: the number of eigenvalues of C greater than 0.7
2. Cumulative summation: the number of ranked eigenvalues of C which explain more than 80% of the variance of C.
3. T-test: the number of eigenvectors of C that have a significant projection onto the observations. Significance tested by means of a t-test.

The results of these tests yield the maximum number of signals that can be considered simultaneously. In the previous studies in which this check has been implemented (Tett et al., 1999, 2001, Stott et al., 2001), the test has been performed off-line from the regression analysis, once, with all possible columns of X* present and at the maximum possible truncation. This is the method employed in this thesis, although there is evidence that it may be sub-optimal, and could be better employed within the regression analysis, rather than as an off-line calculation. Certain combinations which sub-sample signals from X*, can be shown to pass at higher numbers of signals than implied to be possible by the single off-line calculation on all columns of X* (Gareth Jones, personal communication, 2001).

4.2.1.6 Signal-to-noise ratio analysis

AS01 note how OLS estimators are likely to be negatively biased, particularly in their upper bounds, and for weak or poorly defined signals (see their Figure 1 and accompanying text). To address this concern Stott et al. (2001) and Tett et al. (2001) explicitly calculate SNR values, following the approach of North and Stevens (1998). For each signal the SNR value is given by:

$$\left(SNR_j\right)^2 = \frac{m_j}{\kappa}\sum_{i=1}^{\kappa}\frac{X_{ij}^2}{C_{N_{2ii}}} \qquad (21)$$

where m_j is the number of ensemble members used in the calculation of the jth signal, and κ is the truncation. Therefore, calculated SNR values will change with the

truncation being considered. In cases where the signal is pure gaussian WHITE noise the expected value of the SNR will be unity, and SNR values would be approximately F-distributed about this value in a similar manner to consistency tests on the residuals (Tett et al., 2001, see section 4.2.1.3). Therefore, Tett et al. (2001) advocate employing an F-test at the 90% confidence limit based upon these SNRs, to determine whether the signals being considered are likely to be significantly noise contaminated. If they are found to be likely to be significantly noise contaminated, then caution should be placed in inferring firm conclusions based upon OLS estimators, and TLS estimators should preferably be calculated as these are less likely to be systematically negatively biased (Simon Tett, personal communication, 2001). Even in such cases, noise in the signals may effectively render the analysis meaningless under a TLS approach (AS01, see section 4.2.2.3).

4.2.1.7 Applications of the results from OLS regressions and remaining uncertainties

Having implemented a regression algorithm, and checked for consistency of the residuals with realisations of natural internal variability, two potential tests immediately arise. The first test is that of detection: "is the realisation of the signal amplitude in the observations significantly different to that expected by chance?". This can be verified by considering the lower limit of the probable (univariate) distribution at a given confidence interval [generally the 90% level in detection studies (AT99, Hegerl et al., 1996, Stott et al., 2001, Tett et al., 1999, 2001, Barnett et al., 1999)]. If this lower limit is greater than zero then the signal is said to be detected. It should be noted that in this definition detection is solely statistical in nature, it does not *a priori* imply that the signal is important in explaining trends in the atmospheric variable being considered over the detection period.

The second test is one for consistency: "is the model-predicted signal consistent with the observations?". If this is true then the uncertainty limits in the signal amplitude estimates will encompass the value of unity at the prescribed confidence interval

(Stott et al., 2001, Tett et al., 2001). In the single-signal ensemble case, these limits are scaled by $\sqrt{(1+1/m)}$,where m is the number of ensemble members, to account for uncertainty introduced by the finite ensemble size. For a greater number of input signal ensembles, the uncertainty is increased by the product of their individual scaled signal uncertainty estimates (Tett et al., 2001). Stott et al. (2001) show how the univariate confidence limits are distinct from the limits of any multivariate PDF isopleth due to these considerations. A signal can still be detected even if the multivariate PDF isopleth encompasses the origin in the signal dimension, as univariate confidence limits are smaller than the corresponding multivariate limits.

Attribution, as discussed in chapter 1, involves the consideration of all hypothesised causes and plausible combinations of causes until only one remains. Strictly speaking, this yields an infinite population of potential forcing mechanisms to be considered. In practice, it is sensible to limit consideration to a small number of physically plausible mechanisms (AT99). The first step in attribution must be a check for consistency as outlined above. If the signal, or combination of signals, is inconsistent with the observations (i.e. the n-dimensional PDF does not encompass unity in all n dimensions at some critical value) then it cannot be considered to be an adequate representation of the recently observed climate system. The causes of recent climate change are generally attributed to the most parsimonious explanation which does not fail this criteria (Tett et al., 1999, Stott et al., 2001). However, the test in attribution is effectively one of non-rejection of a null hypothesis of consistency with the observations. Due to the asymmetric nature of the statistical system, rejection of the alternative hypothesis at a confidence level P does not imply acceptance of the null at a confidence level of 1-P (Levine and Berliner, 1999). Any such claims are likely to be significantly overestimating the significance of the statistical acceptance of the null.

Considerable uncertainties can remain in results from detection studies employing an OLS approach. The primary caveat that must be applied is that the algorithm does not explicitly take into account noise in the estimated fingerprints. Using ensemble

average fields to estimate the columns of X in the OLS approach will give potentially significant non-zero noise in the signals if the ensemble size is small, or the signal is weak. This is likely to impact most greatly upon the upper limits of the PDF distribution, which are underestimated using an OLS approach (AT99). AS01 illustrate this by showing how, in reality, OLS involves estimating the ratio of the amplitudes of the observations and the model-simulated response. If the signal incorporates noise, then the denominator (model amplitude) could be overestimated introducing an underestimate in the signal strength estimator, especially for weak signals where noise can become dominant. For a complete explanation of this bias the reader is directed to Figure 1 of AS01, together with accompanying text. The TLS algorithm detailed in the next sub-section (AS01) overcomes this problem. A further possible avenue to constrain uncertainty is to increase ensemble size such that the "distinguishing fingerprints" contain a purer realisation of the true signal (AT99, AS01, Barnett et al., 1999).

Secondly, there is uncertainty associated with the pre-whitening factor, P, in that it is an estimate based upon a finite sample from control and/or intra-ensemble variability. If another choice of control run were used, then the expectation is that P, and hence the results of the regression, would be different to some extent. Assuming the model control variability is adequate, i.e. not exhibiting systematic biases, this can be considered to be essentially a random error.

A further, potentially more worrying, problem is if the signal were associated with a change in the higher order statistics of the modes onto which it projected, due to climate non-linearities (see Palmer, 1999 for a discussion of this), then this would not be included in the regression. AS01 address the concern by use of a hypothetical example based upon a simple Lorenz attractor example. They illustrate how over long time periods, greater than that of the exponential error growth timescale of the system, such a bimodal chaotic system is indistinguishable from a normally distributed system, as expected from a consideration of the Central Limit Theorem. In the example given (AS01), the response is also not in the direction of the forcing

applied, but this is not important, assuming that the model adequately captures those factors important in explaining the system response. In view of the fact that detection studies are essentially tests of model adequacy, even if the model does not achieve this it should not be a problem, as one would reject the model as a plausible explanation of climate changes, so it can only lead to conservative errors. The real world system is much more chaotic than the simple example given in AS01, and can be viewed as a multiple potential well (with multiple rather than two preferred states) system (Corti et al., 1999). However, the same considerations should yield the same conclusion as that for the idealised conceptual example of AS01, although it is not possible to rigorously test this. The climate system is, therefore, likely to be able to have OLS regression tools applied to it if the timescales considered are adequately long, even though on shorter timescales the system exhibits distinct non-linear chaotic characteristics (Corti et al., 1999, for example).

If the model exhibits systematic biases in its estimate of noise properties, then this can influence the results. There are two conceivable systematic biases: an overall deficiency in the estimate of natural variability; and error in the models ability to capture specific modes of variability, especially important when either the modelled or true signal projects strongly onto any such mode (AS01). If the overall variability estimate from the control is incorrectly diagnosed, then this will lead to obvious complications in that the confidence limits both for detection and consistency checks will be inaccurate (Allen et al., 2001). If it is grossly underestimated then the consistency test on the residuals will fail, whereas if it is overestimated then the uncertainty limits will be overestimated. Therefore, in terms of detection this can only ever lead to a conservative error. If a particular real world climate mode is either not included in, or mis-represented by, the control, then there is a potential cost involved in the pre-whitening (optimisation) operator. If the model-predicted signal(s) project(s) strongly onto this mode, then this component will either not be considered, or be incorrectly weighted in the regression algorithm, with obvious implications. This can only ever be conservative, as either the searched-for signal will be sub-optimal, or the consistency check on residuals will identify that the

residuals from the regression are unrealistic, if the mode is grossly misrepresented in the control realisation. Gillett et al. (2000a) show how deficiencies in the model simulations of the Arctic Oscillation do not affect the principal results of previous OLS detection studies (Tett et al., 1999) considering the near-surface temperature record for HadCM2.

4.2.2 Total Least Squares regression

4.2.2.1 Accounting for the presence of noise in the signal estimates

Under OLS regression, an ad hoc scaling correction is applied to the results, to account for the likely presence of noise in the signals due to the finite ensemble size available. Total Least Squares (TLS) regression advances this to implicitly account for noise in the columns of X, the model predicted signals, as well as y, the observations, in the regression algorithm. AS01 illustrate how this is, mathematically speaking, the correct methodology for ascertaining amplitude estimates in the presence of uncertainty, through a hypothetical example (see their Figure 1 and accompanying text for more details).

Returning to the basic underlying OLS regression model (1), this is replaced in the TLS algorithm by:

$$y = \sum_{i=1}^{m} \left(X_i - \nu_{x_i} \right) \beta_i + \nu_y \tag{22}$$

Hence the sole underlying difference in the TLS regression approach is that an extra noise-term has been added to the independent predictor variables in the regression. An assumption in currently implemented TLS algorithms is that the noise variance in both the individual columns of X (the signals, which are scaled if necessary to account for ensemble averaging effects), and the vector y (the observations), arises solely from natural variability ($\nu_x = \nu_y$). This natural variability is estimated from a model control run which, as for the OLS approach, is assumed to adequately

represent the leading κ modes of variability within which the regression is implemented (AS01). The assumption that the noise in each column of X is identical to that of y, and hence real-world variability, yields (c.f. (3)):

$$\varepsilon\left(P\nu_{x_i}\nu_{x_i}^{T}P^{T}\right)=I_{\kappa} \tag{23}$$

$$\varepsilon\left(P\nu_{y}\nu_{y}^{T}P^{T}\right)=I_{\kappa} \tag{24}$$

Only if these equations hold true, and both pre-whitened observed and modelled fields appear to be WHITE, can TLS tools be applied to the climate change detection problem. Therefore, if X, or y, contain non-negligible error terms, expectations are for potentially highly biased estimators to result under a TLS approach. If any such errors can be quantified in a meaningful manner then they could in theory be incorporated, at least to some extent, as an additional covariance matrix within the regression (AS01). Following the analysis in chapter 2, it is likely that the HadRT record used in this thesis contains at least some residual errors, not all of which are likely to be negligible. Nor are they likely to be easily accurately estimated in any quantitative manner. Therefore, TLS tools may not be applicable to at least some upper air temperature diagnostics based upon HadRT records. If the columns of X are ensemble average values, then the expected noise variance can simply be scaled up to enable a direct comparison with y, for which, at least in applications to date, there can only ever be one member. Identical scaling is then applied to the results of the TLS regression (AS01).

As in the OLS solution, consideration in the TLS approach is limited to the leading κ modes of variability, as estimated from a finite section of model control. Because there is no longer any distinction between the columns of X, and y, under a TLS approach, they can be treated simultaneously within the regression. If $m' \equiv m+1$ then the $m' \times \kappa$ matrix:

$$Z \equiv \left[PX, Py\right] \tag{25}$$

is defined as the pre-whitened observed values of X and y where P, the pre-whitening operator, is defined in exactly the same manner as for the OLS solution (3). It should

be noted that, although these values will be pre-whitened, they will still be noise contaminated (AS01).

The multi-linear regression model proposed by AS01 implies that there must be a value Z^{true}, for which the columns of Z are linearly related. Therefore, a solution-set must exist, whereby there is a combination of coefficients on Z which is an internally consistent explanation, and hence perfectly recreates the observations such that (22) becomes:

$$Z^{true}V = (Z - Y)V = 0 \tag{26}$$

where Z^{true} is the true, non-contaminated model signals and observations, V is a vector of coefficients of rank m′ (the multipliers of each column of Z), and Y is the $m' \times \kappa$ matrix of true (pre-whitened) noise contamination (residuals) in the model and observed variables (AS01). As all the elements of Y are normally distributed with unit variance (as the input has been pre-whitened), the maximum likelihood estimator of V, $\widetilde{V}$, the best-guess signal and observed coefficients in the TLS regression model, is found by maximising:

$$L = const - \frac{1}{2} tr\left(\widetilde{Y}^T \widetilde{Y}\right) \tag{27}$$

where L is the likelihood estimator, const = a constant (an arbitrary value to enable the derivative of L to be realised), $\widetilde{Y} = Z - \widetilde{Z}^{TRUE}$ (the best estimate residuals), and $\widetilde{Z}^{TRUE}\widetilde{V} = 0$ (satisfying (26)) (AS01).

The rows of the residuals, $\widetilde{Y}$, are uncorrelated with the coefficient estimates, $\widetilde{V}$, and therefore the solution to (27) is equivalent to minimising:

$$r^2 = \widetilde{V}^T \widetilde{Y}^T \widetilde{Y} \widetilde{V} \tag{28}$$

The trivial solution of $\widetilde{V} = 0$ must be avoided in any approach (AS01). To avoid bias in any constraint placed upon the result of (28), the standard normalisation of $\widetilde{V}^T \widetilde{V} = 1$ is employed (AS01). The standard normalisation can only be employed as there is effectively no distinction between individual columns of Z under the TLS approach. Applying this constraint to (28) yields:

$$r^2\left(\widetilde{V}\right) = \widetilde{V}^T Z^T Z \widetilde{V} + \lambda^2\left(1 - \widetilde{V}^T \widetilde{V}\right) \tag{29}$$

where λ^2 is a Lagrange multiplier (AS01).

Minimising r^2 in (29) can be shown to be equivalent in geometric terms to finding the m-dimensional plane in an m'-dimensional space such that the sum of squared distances from the plane to the κ points defined by the rows of Z is minimised (AS01). Hence the terminology of a Total Least Squares solution.

Differentiation of (29) with respect to $\widetilde{V}$ gives an eigen-equation yielding the stationary points of r^2 (AS01), whereby:

$$\frac{\partial\left(r^2\right)}{\partial\left(\widetilde{V}\right)} = Z^T Z \widetilde{V} - \lambda^2 \widetilde{V} = 0 \tag{30}$$

and the curvature matrix is given by:

$$\frac{1}{2}\frac{\partial^2\left(r^2\right)}{\partial\left(\widetilde{V}\right)^2} = Z^T Z - \lambda^2 I \tag{31}$$

r^2 is therefore minimised by minimising the Lagrange multiplier term, such that $\lambda^2=\lambda^2_{min}$, which from (31) is gained by finding the smallest eigenvalue of Z^TZ, with corresponding eigenvector $\widetilde{V}$ (being the vector normal to the best fit m-dimensional plane) (AS01). This can be shown to be the best (lowest variance) unbiased estimator in geometric terms. Implementation is achieved by consideration of the singular value decomposition $\mathbf{Z=Q\Lambda R^T}$ such that, after sorting, $\widetilde{V} = R_{m'}$, the eigenvector corresponding to the smallest eigenvalue, Λ (AS01).

The m'th element of the resulting solution vector, $\widetilde{V}$, corresponds to the best-fit scaling parameter on the observation vector, y. AS01 therefore suggest that for analysis purposes, the coefficients are translated into the more familiar pattern-amplitudes yielded in the OLS solution, by finding the ratios $\widetilde{\beta}_i = \widetilde{V}_i / \widetilde{V}_{m'}$. The actual values of V themselves are physically meaningless being geometrically a representation of the angles of the plane in the m-dimensional phase space described

by the TLS solution. At the minimum gained from (31) the distribution on these $\tilde{\beta}$ estimates, assuming the noise estimate from the model control is accurate (or at the very least adequate), is approximately χ^2:

$$r_{\min}^2 = \lambda_{\min}^2 \approx \chi_{\kappa-m}^2 \tag{32}$$

This distributional assumption holds true provided that $\kappa >> m$. This relationship provides for the construction of a consistency test on the residuals that is directly analogous to that of the OLS approach (15) (AS01).

Analysis of uncertainty in the best-guess estimate of the signal amplitudes, $\tilde{\beta}$, is more complicated than that employed in the OLS approach (AS01). Returning to the singular value decomposition $Z=Q\Lambda R^T$, the diagonal matrix Λ^2 contains the ranked eigenvalues of Z^TZ, and the columns of R contain R_i, the corresponding eigenvectors. Hence equation (31) can be re-written as:

$$\frac{1}{2}\frac{\partial^2\left(r^2\right)}{\partial\left(\tilde{V}\right)^2} = R\left(\Lambda^2 - \lambda_{\min}^2\right)R^T \tag{33}$$

In the limit of a high SNR, it is reasonable to treat the pseudo-inverse of (33) as a standard covariance matrix on $\tilde{V}$, whilst in the presence of weak signals this may not hold true (AS01). Because the normalisation constraint on the coefficients, $\tilde{V}$,(29) introduces a non-linearity into the system, the merit function r^2 is not quadratic. Realistic PDF isopleths can be derived by plotting surfaces of V where $\Delta r^2(V) = r^2(V) - r_{\min}^2$ evaluates to a constant value. As for the OLS model the solution evaluates, at least approximately, to a chi-squared distribution (AS01):

$$\Delta r^2(V) = V^T R\left(\Lambda^2 - \lambda_{\min}^2 I\right)R^T V \approx \chi_m^2 \tag{34}$$

Subsequently, a confidence interval is chosen, and those vectors of V for which $\Delta r^2(V)$ is equal to the critical χ_m^2 value are evaluated to form an isopleth by evaluating the relevant values for the LHS of (34).

AS01 show how such a mapping is computationally realised by firstly defining a set of points on an m-dimensional sphere of radius $\sqrt{\Delta r_{crit}^2}$, the critical χ_m^2 distribution:

$$\sum_{i=1}^{m} a_i^2 = \Delta r_{crit}^2 \tag{35}$$

For each of the points on the sphere, a value b_i is then computed such that:

$$b_i = \frac{a_i}{\sqrt{\lambda_i^2 - \lambda_{\min}^2}} \tag{36}$$

If the b_i provide the weights on eigenvectors 1 to m in R used to generate V, then (34) must be satisfied. The weight on $V_{m'}$ is provided by the normalisation constraint:

$$\sum_{i=1}^{m'} b_i^2 = 1 \tag{37}$$

If Δr_{crit}^2 is too large, $b_{m'}$ will be calculated to be either zero or imaginary, and the confidence interval will be boundless in at least one dimension. This implies that the signal coefficients, V, are unconstrained to rotations through 360° in some plane at the specified confidence interval (AS01). This is likely to be the case when the signals are either weak and poorly constrained by the model, or highly degenerate. It should be noted that this approach will tend to yield isopleths which are far from ellipsoidal. Therefore, although some signals in a multiple input-signal case may be unconstrained, other signals may still be well constrained under the TLS approach (AS01).

4.2.2.2 Accounting for uncertainty in the control estimate of natural variability

Up until this point in the TLS algorithm an implicit assumption has been that natural internal variability is perfectly known, based upon the finite section of model control used to define the pre-whitening operator, P. As in the OLS approach, there is reason to believe that this is not the case. Therefore, two separate segments of control should also be used for optimisation and significance testing respectively in the TLS approach (AS01). AS01 further note that in taking into account noise in all variables, the estimation algorithm has become non-linear. Therefore, it is no longer valid to realise a series of $\tilde{\beta}$ -like estimates from the columns of $\hat{Y}_2$, the independent control

estimate, as was the case in the OLS algorithm (12), to provide a confidence interval. AS01 state that each λ_i^2 represents the signal-to-noise ratio in the corresponding pair of singular vectors of Z:

$$\lambda_i^2 = \frac{Q_i^T Z Z^T Q_i}{Q_i^T P_1^T \hat{C}_N P^T Q_i} \tag{38}$$

Where Q_i are the left singular vectors of the singular value decomposition of Z. In this equation the denominator will be unity if the μ columns of $\hat{Y}_1$, the first control segment used to estimate $\hat{C}_N$, have also been employed in the derivation of P, the pre-whitening operator. As in the OLS algorithm, relying upon these λ estimates for uncertainty analysis may be misleading if $\hat{Y}_1 \hat{Y}_1^T \left(\hat{C}_N\right)$ is rank deficient, which in the limit of a very long control run it invariably will be. Poorly-realised state-space directions will be given artificially high weight by the pre-whitening operator since $P\hat{C}_N P^T = I$. This will artificially inflate the differences between eigenvalues and, therefore, reduce estimated uncertainties (AS01).

As for the OLS approach, the solution is to replace the segment of control, $\hat{C}_N$, used for optimisation (pre-whitening) with an independent realisation of natural variability, $\hat{C}_{N2}$, when realising confidence intervals (AS01):

$$\hat{\lambda}_i^2 = \frac{Q_i^T Z Z^T Q_i}{Q_i^T P \hat{C}_{N2} P^T Q_i} \tag{39}$$

As $\hat{C}_{N2}$ is independent of P, any systematic bias in P will apply to both the numerator and denominator in (39), whereas it will only apply to the numerator in (38) (AS01). This avoids any systematic bias in the calculation of the confidence region, whilst having no effect upon the best guess estimates $\widetilde{\beta}$, and $\widetilde{V}$. As is the case in the OLS algorithm, the check for residual consistency must be changed such that:

$$\hat{r}_{\min}^2 = \hat{\lambda}_{\min}^2 \approx (\kappa - m) F_{(\kappa-m),\nu_2} \tag{40}$$

and (34) becomes:

$$\Delta\hat{r}^2(V) = V^T R\left(\hat{\Lambda}^2 - \hat{\lambda}^2_{\min} I\right) R^T V \approx mF_{m,\nu_2} \quad (41)$$

Again, confidence limits are realised by calculating values for the LHS of (41) for critical values of the appropriate F-distribution.

4.2.2.3 Remaining problems

In TLS regression there is an important caveat to consider, whereby if the true signal amplitude is weak or close to zero, then the confidence bounds on the estimator, $\tilde{\beta}$, will tend to become infinitely large (AS01). This may appear counter-intuitive at first, however in such a case there is no reason for $\tilde{\beta}$ to prefer any orientation in the β phase space, as both model and observed signal amplitudes will tend to zero. Orientation in this case is essentially arbitrary in the TLS solution, and the ratios $\tilde{\beta}_i = \tilde{V}_i / \tilde{V}_{m'}$ can, therefore, take any value. Thus, a near infinite pattern-amplitude implies that, because a TLS approach incorporates noise in the signals, the true response-pattern may be close to zero in all directions. For stronger signals the main difference in results compared to an OLS approach will be both a higher best-guess signal amplitude, and substantially increased upper bounds to the uncertainty distribution (see AS01 for a theoretical justification). If the upper bound is sufficiently uncertain, then the results will essentially be meaningless. As the orientation of the plane $\left(\tilde{V}\right)$ passes through 90° in any signal direction, the value of the amplitude estimate $\left(\tilde{\beta}\right)$ goes from +∞ to -∞, and then progressively smaller negative finite numbers. There is little effect upon the lower bounds, which are the important component in detection exercises, although care must be taken within any analysis to avoid confusion when the upper-bounds are unconstrained, and therefore negative (Peter Stott, personal communication, 2001).

AS01 contend that an argument ensues as to whether the extra mathematical complication of the TLS approach is justified above and beyond the OLS algorithm, results of which are also easier to interpret. Two points are important for

consideration as to why TLS should, at least in some cases, be considered. Firstly, attribution requires the rejection of all other plausible signals (or combinations thereof) until only one remains. Therefore, as OLS regression systematically under-estimates the upper bounds of the uncertainty range, in the limit of an infinite ensemble, reliance on OLS may lead to an ambiguous claim of attribution. Secondly, Allen et al. (2000, 2001), have advocated the use of regression to constrain model projections of future climate change, and illustrated this with regard to global mean near-surface temperatures. If, as is the case in OLS regression, the upper bounds of the amplitude uncertainty range are systematically and significantly underestimated, at least for the small ensemble sizes normally available, then the use of such an algorithm may underestimate the upper confidence limit on the constrained projections. Such an approach could, therefore, lull end users of the data into a false sense of security regarding the likely amplitude of a worst-case scenario prediction.

4.3 Other multi-variate detection approaches

Numerous other approaches have been advocated in climate change detection and attribution studies. Hasselmann (1997) provided the motivation for the development of the OLS solution detailed in the previous section (AT99), and in turn the motivation for Hasselmann (1997) came from previous studies. Here a number of these individual approaches are very briefly described, and where possible presented in terms of the OLS regression methodology defined in the previous section. The similarities are stressed to illustrate that all approaches originate from the same population of potential approaches, and how the regression approach is indeed the most powerful of these (Levine and Berliner, 1999).

The pattern correlation approaches of Santer et al. (1993, 1995) can be shown to be sub-optimal versions of the regression algorithm (Mitchell and Karoly et al., 2001). Considering a single signal in the OLS regression approach, the solution is:

$$\widetilde{\beta} = \frac{X_1^T C_N^{-1} y}{X_1^T C_N^{-1} X_1} \qquad \text{(compare to (4))} \tag{42}$$

The uncentred correlation statistic C(t) (Santer et al., 1995) can be written in similar terms, adopting array notation, and incorporating an extra term in the denominator (Mitchell and Karoly et al., 2001):

$$C(t) = \frac{X_1^T y}{X_1^T X_1} \equiv \frac{X_1^T I y}{X_1^T I X_1} \tag{43}$$

In both cases, X_1 is a column vector containing a single model signal realisation. So C(t) is simply the unoptimised version of the regression algorithm, whereby the n×n covariance matrix is an identity matrix. Similarly, the centred statistic R(t) can be denoted in matrix notation (Mitchell and Karoly et al., 2001):

$$R(t) = \frac{X_1^T (I-U) y}{\left[\left(X_1^T (I-U) X_1 \right)^{1/2} \left(y^T (I-U) y \right)^{1/2} \right]} \tag{44}$$

Here U is an n×n matrix with elements $u_{i,j}=1/n$ which remove the spatial means from the observed and modelled series. Both measures (43 and 44) are constrained to lie in the interval –1 to 1 and, therefore, cannot be employed directly to estimate the signal amplitudes.

Pattern correlation approaches are limited to considering a single signal, or a single combination of signals with pre-determined weights. This means they are less powerful than those optimal regression techniques that can consider multiple signals simultaneously. For both the R(t) and C(t) statistics, no effort is made to optimise the data by reference to the SNR values and, therefore, the approach is likely to have a large cost (be highly inefficient) when attempting to detect an emerging signal in a noisy background (AT99, AS01). An advantage is that it does not depend on the model providing an adequate realisation of natural climate variability in defining the signal, and, therefore, should be seen as complementary to optimal approaches. If the model control estimate is grossly inadequate then the pattern correlation technique may be more efficient than the optimal regression algorithms discussed in section 4.2. However, for both these pattern correlation statistics, significance is still estimated with reference to model control data and, therefore, a statistical cost function remains if the control is inadequate at the scales considered. Previous correlation approaches which have considered near-surface temperatures have

considered the raw fields (Santer et al., 1995, Wigley et al., 1998). There is evidence that at least HadCM2 underestimates natural variability at scales below 2000Km (Stott and Tett, 1998). Therefore, the confidence limits may be systematically underestimated in applications of the pattern correlation approach to date, increasing the chances of a Type I statistical error arising. This could be circumvented by truncating the input fields in space, but has not been attempted to date.

Bell (1982) was the first to advocate a formal optimal detection approach, taking into account knowledge of the climate system covariance. In the algorithm proposed by Bell (1982), each term is given a weight such that the SNR is maximal under the constraint of the weights summing to unity $(w = \alpha \hat{C}_n^{-1} X)$, with a constant factor multiplier, α. It is assumed under this approach that the weighted observations are a linear combination of the weighted signal and noise, with the signal assumed to be of both the correct pattern and amplitude; $y = X + \nu$. There is strong evidence that neither of these assumptions is true, especially in the case of weak signals (AS01) and, therefore, that this approach is highly sub-optimal, and could lead to erroneous conclusions. The algorithm can only consider a single signal at a time and evaluates the statistic:

$$\frac{y}{\sqrt{\nu^2}} \tag{45}$$

If this statistic is above a certain threshold, the signal is detected at the specified confidence interval.

Hasselmann (1997) details an optimal detection methodology whereby a series of fixed signals are derived from the difference between a present day and a late 21st Century realisation of climate change. Fingerprints are constructed by stipulating an orthogonality constraint in the multiple signal case before proceeding to detection in applications of this methodology (Hegerl et al., 1997), which can lead to problems in interpretation (AT99). In the limit of noisy signals, this is likely to have a cost function appended. However, the use of a 21st Century difference field mitigates this effect, as the derived signals are effectively noise free. The methodology implicitly

assumes that the climate response pattern is not transient, being solely a linear multiplier of the searched-for signal response(s), which is (are) a (some) fixed pattern(s). Optimisation is carried out with respect to the spatial covariance matrix as estimated from a model control run. Pattern congruence tests are then used to derive the detection statistics. The statistic couched in terms of the regression algorithms discussed in section 4.2 can be shown for a single signal to be:

$$\hat{X} = \frac{X^T \hat{C}_n^{-1} y}{X^T \hat{C}_n^{-1} X} X \tag{46}$$

where $\hat{X}$ on the LHS is the best guess signal amplitude. However, the absolute amplitude of the signal in the observations cannot be calculated as this statistic incorporates information both on the signal pattern and its amplitude through the extra term in the RHS of (46) when compared to (4). The statistic can be calculated for consecutive observational dataset time slices over a moving window, and trends evaluated and compared to those arising by chance from an independent control run to assess the significance.

Finally, North et al., (1995) detail a methodology that undertakes optimisation in both space and time to elicit the size of the signal in noise contaminated data. The signal is optimised by integrating over both space and time incorporating a kernel, yielding a signal based upon the filtered data. To date only single signals have been considered, although extension to a multiple signal approach is straightforward (Hegerl and North, 1997). The kernel is defined such that the mean square error between the signal estimate and original signal is minimised. Optimisation is therefore implemented with a no bias constraint. Optimisation occurs in space-time EOF space, where the EOFs represent the orthogonal components of the space-time lagged covariance matrix. The final product can be shown to be directly comparable to Hasselmann (1997) if data are discrete, which they invariably are (Hegerl and North, 1997). A by-product of this methodology is that SNRs can be explicitly calculated. Stott et al. (2001) show how it is possible to employ such an approach using OLS regression tools.

4.4 Conclusions

All detection studies to date that consider some form of pattern similarity can be shown to be directly related (Hegerl and North, 1997, Levine and Berliner, 1999, AT99, Mitchell and Karoly et al., 2001). The differences between these methodologies are both in terms of their complexity, and the tools that can be applied to the output. In this thesis the preferred statistic is a form of least squares regression as outlined in section 4.2, either Ordinary Least Squares (OLS, 4.2.1), or Total Least Squares (TLS, 4.2.2). Regression has advantages over the alternative methods in terms of being able to estimate signal amplitudes directly and, therefore, make meaningful statements about both signal detection and consistency (AT99, Tett et al., 2001). It is also more flexible than other approaches in terms of both the number of signals able to be considered, and the fact that these can be easily recombined to yield their individual components. Importantly, both OLS and TLS regression algorithms provide for a test on the residuals that should flag when the results are (grossly) unrealistic. The regression approach can also potentially be used to constrain future projections in a meaningful manner (Allen et al., 2000, 2001). Levine and Berliner (1999) further show how regression is the uniformly most powerful statistic for detection purposes under a classical frequentist statistical approach.

The regression approach can be shown to be a special case Bayesian statistic, whereby the prior probability is uniform (Allen et al., 2001). To date there has only been one published study which has used a Bayesian framework in a multi-variate climate change detection setting (Berliner et al., 2000), and even then in a highly idealised exercise. Efforts are underway elsewhere to apply Bayesian analysis (Gareth Jones, personal communication, 2001). Although it is recognised that this is a likely and desirable advancement in future detection studies, it has not been covered in any detail in this chapter, as it is not employed in this thesis.

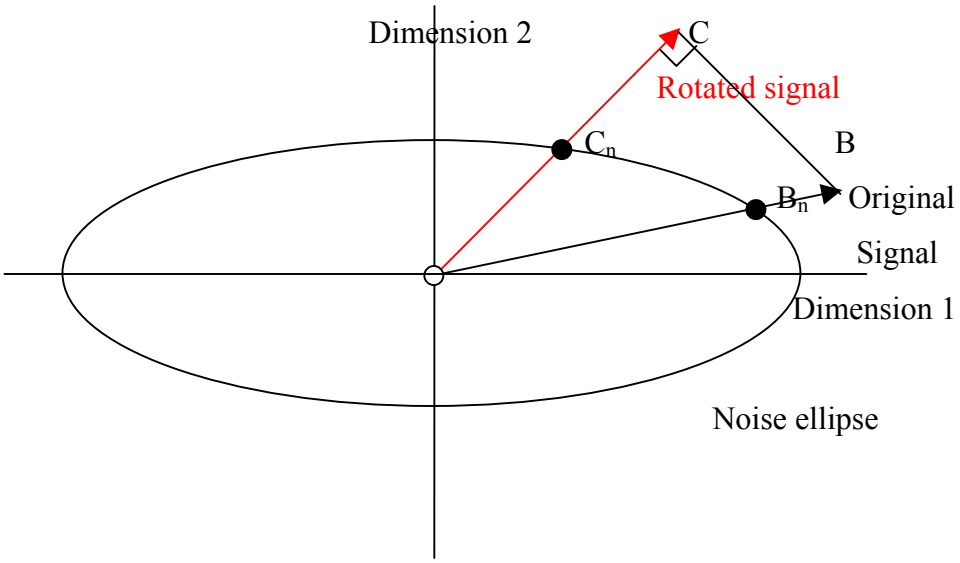

Figure 4.1 (after Hasselmann, 1976, and Mitchell and Karoly et al., 2001) Figure showing a simplified rationale for rotation of signals in climate change detection studies. The ellipse shows the 95% distribution of natural variability in the parameter in the simple two dimensional space. The black arrow 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).