Scale-Dependent Detection of Climate Change

PETER A. STOTT AND SIMON F. B. TETT

Hadley Centre for Climate Prediction and Research, U.K. Meteorological Office, Bracknell, Berkshire, United Kingdom

(Manuscript received 7 July 1997, in final form 4 February 1998)

ABSTRACT

Spatially and temporally dependent fingerprint patterns of near-surface temperature change are derived from transient climate simulations of the second Hadley Centre coupled ocean–atmosphere GCM (HADCM2). Trends in near-surface temperature are calculated from simulations in which HADCM2 is forced with historical increases in greenhouse gases only and with both greenhouse gases and anthropogenic sulfur emissions. For each response an ensemble of four simulations is carried out. An estimate of the natural internal variability of the ocean–atmosphere system is taken from a long multicentury control run of HADCM2.

The aim of the study is to investigate the spatial and temporal scales on which it is possible to detect a significant change in climate. Temporal scales are determined by taking temperature trends over 10, 30, and 50 yr using annual mean data, and spatial scales are defined by projecting these trends onto spherical harmonics.

Each fingerprint pattern is projected onto the recent observed pattern to give a scalar detection variable. This is compared with the distribution expected from natural variability, estimated by projecting the fingerprint pattern onto a distribution of patterns taken from the control run. Detection is claimed if the detection variable is greater than the 95th percentile of the distribution expected from natural variability. The results show that climate change can be detected on the global mean scale for 30- and 50-yr trends but not for 10-yr trends, assuming that the model's estimate of variability is correct. At subglobal scales, climate change can be detected only for 50-yr trends and only for large spatial scales (greater than 5000 km).

Patterns of near-surface temperature trends for the 50 yr up to 1995 from the simulation that includes only greenhouse gas forcing are inconsistent with the observed patterns at small spatial scales (less than 2000 km). In contrast, patterns of temperature trends for the simulation that includes both greenhouse gas and sulfate forcing are consistent with the observed patterns at all spatial scales.

The possible limits to future detectability are investigated by taking one member of each ensemble to represent the observations and other members of the ensemble to represent model realizations of future temperature trends. The results show that for trends to 1995 the probability of detection is greatest at spatial scales greater than 5000 km. As the future signal of climate change becomes larger relative to the noise of natural variability, detection becomes very likely at all spatial scales by the middle of the next century.

The model underestimates climate variability as seen in the observations at spatial scales less than 2000 km. Therefore, some caution must be exercised when interpreting model-based detection results that include a contribution of small spatial scales to the climate change fingerprint.

1. Introduction

A number of recent studies have shown evidence for a discernible human influence on climate (e.g., Mitchell et al. 1995; Hegerl et al. 1996; Santer et al. 1996; Tett et al. 1996; Hegerl et al. 1997). Climate change detection and attribution studies involve the identification of a model-predicted climate change signal in the observations and the demonstration that the strength of the signal is greater than expected due to natural variability. Validation of model simulations against observations in this way raises confidence in model-based predictions of future climate change. The aim of this study is to assess the present and future feasibility of regional climate change prediction by investigating the spatial and temporal scales on which it is possible to detect a significant change in climate.

In this study, modeled patterns of climate change are taken from transient experiments in which the second Hadley Centre coupled ocean–atmosphere general circulation model (HADCM2) (Johns et al. 1997) is forced with historical increases in greenhouse gases and sulfate aerosols. Mitchell et al. (1995) showed that the inclusion of sulfate aerosols in addition to greenhouse gases significantly improved the agreement with observed global mean and large-scale patterns of near-surface temperature in recent decades. Several other recent studies have compared observed and modeled patterns of climate change from coupled model simulations that incorporate both greenhouse gases and anthropogenic sulfate aerosols. Santer et al. (1995) showed that the most recent 50-yr near-surface temperature trends in a centered cor-

Corresponding author address: Dr. Peter A. Stott, Hadley Centre for Climate Prediction and Research, U.K. Meteorological Office, London Road, Bracknell, Berkshire RG12 2SY, United Kingdom.
E-mail: pastott@meto.gov.uk

Unauthenticated | Downloaded 06/13/21 11:50 PM UTC

relation statistic for the July–August and September–November signals were statistically significant relative to model-based estimates of internal variability. Hegerl et al. (1997) used an optimal fingerprint method applied to near-surface temperature trends and showed that the most recent trends represented a significant climate change. Vertical patterns of zonally averaged temperature change have been investigated by Santer et al. (1996) and Tett et al. (1996), where the use of centered correlation statistics also showed a significant degree of pattern similarity between models and observations.

The methods used in all these studies measured the degree of similarity between modeled and observed temperature changes at all spatial scales, but only at the largest spatial scales did they see significant agreements between models and observations. As we go to shorter spatial scales it is likely to become more difficult to detect the signal of climate change from the noise of natural variability. We aim to address the following questions:

1) On what spatiotemporal scales does HADCM2 currently have skill at modeling observed patterns of near-surface temperature change?
   - If the level of agreement between modeled and observed patterns of change is significantly different from zero (i.e., not attributable to changes expected from natural variability), then climate change detection can be claimed.
   - We also investigate if either of the simulations (greenhouse gas alone, greenhouse gas and aerosol forcing) can be shown to be inconsistent with the observations.
2) On what scales is climate change detectable? We aim to determine how the probability of detection varies with temporal and spatial scales at present and in the future, assuming that our model simulations perfectly represent the observed climate change to within the limits of natural variability.
3) On what scales does the model adequately represent the internal variability of the climate system? This is important since detection claims depend crucially on the reliability of the model's estimate of natural variability.

The strategy we apply is, as in Hegerl et al. (1996), to use patterns of near-surface temperature trends simulated by models, to represent the expected fingerprint of anthropogenic climate change. However, whereas Hegerl et al. (1996) identified the expected climate change by calculating the global projection of the observations onto a statistically optimized fingerprint, here fingerprint patterns and observations are defined at a range of spatial as well as temporal scales.

Temporal scales are determined by taking temperature trends over N yr using annual mean data. Spatial scales are defined by projecting these trends onto spherical harmonics. The aim here is not to develop and apply a detection algorithm that is optimal such as the method described by Hegerl et al. (1996). Instead, we aim to determine those spatial and temporal scales on which it is possible to detect a significant change in climate. Spherical harmonics are the appropriate coordinate system to use here, since they have a natural interpretation in terms of spatial scales. Empirical orthogonal functions on the other hand, onto which Hegerl et al. (1996) projected selected climate change signals, are empirical in nature and do not have natural a priori interpretations in terms of spatial scales. To the extent that the use of spherical harmonics is less than optimal, the method applied in this study may be conservative in that it fails to claim detection when a more optimal method would. There is further discussion of the technique of projecting climatic fields onto spherical harmonics to detect forced climate signals by North et al. (1995), North and Kim (1995), Kim and North (1993), who applied optimal filters to surface temperature fields obtained from an energy-balance model and by North et al. (1992) and Leung and North (1991) who investigated the response of a general circulation model with idealized boundary conditions and forcings.

The structure of this paper is as follows. Section 2 describes the methodologies that we apply, the results are presented in section 3, and the conclusions in section 4. Sections 2 and 3 are each divided into subsections that treat in turn the scales on which the model has skill, the scales on which climate change is detectable in principle, and the scales on which the observed variability is adequately represented by the model.

2. Methodology

Patterns of near-surface temperature trends from simulations of HADCM2 and from observations are projected onto spherical harmonics. Trends are calculated over 10, 30, and 50 yr from transient experiments in which HADCM2 is forced with greenhouse gases only (G) and greenhouse gases and sulfate aerosols (GS). Near-surface temperature trends are also taken from a 1000-yr control run (CTL) of HADCM2 in which greenhouse gases are held constant to represent natural variability (Tett et al. 1997).

The G and GS experiments both consist of an ensemble of four integrations, each of which has identical forcing but has different initial conditions taken 150 yr apart from the 1000-yr control run, to represent the possible spread in forced model trajectories due to natural variability. Both G and GS use estimated forcing perturbations starting in 1860. Until 1990, G is forced with equivalent CO_2 changes that mimic the change of forcing due to all greenhouse gases relative to preindustrial levels. From 1990, CO_2 forcing increases by 1% yr^{-1} (compound) up to 2100, at which time the increase in forcing relative to 1990 (6.5 W m^{-2}) is close to that of the Intergovernmental Panel on Climate Change (IPCC) scenario IS92a (Houghton et al. 1992). For GS, an additional negative forcing is applied by an increase in surface albedo to represent the direct radiative forcing

Unauthenticated | Downloaded 06/13/21 11:50 PM UTC

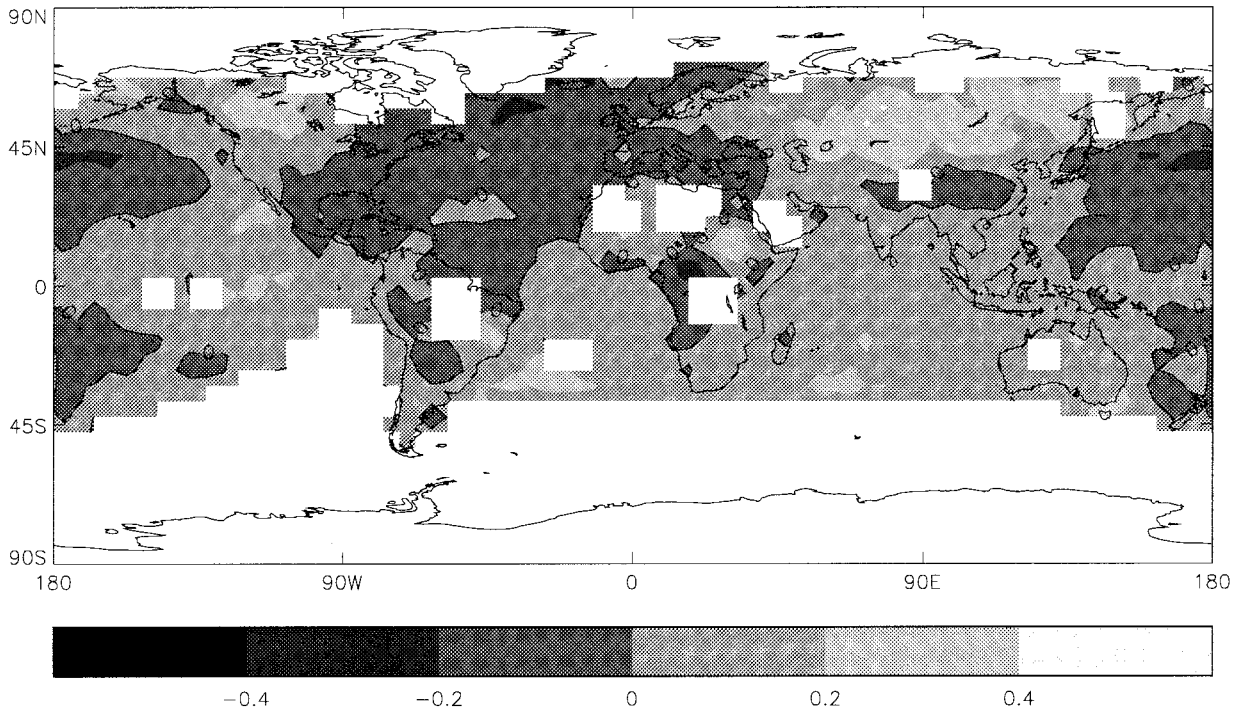

FIG. 1. Observed 50-yr trends to 1995 (K decade^{-1}).

due to anthropogenic sulfate aerosol. The pattern of aerosol loading to 1990 is based on an estimate of sulfate emissions in the 1980s and is scaled by global and decadal mean industrial sulfate emissions to give a geographical distribution of sulfate aerosol from 1860. From 1990 to 2050 the sulfate loading pattern is interpolated between 1990 and 2050 values, where the 2050 values are based on a sulfur-cycle model (Langner and Rodhe 1991). After 2050, the sulfate loading pattern is held constant and scaled by global mean emissions, although the total aerosol forcing changes little after 2050. Further details of the forcing and experimental design for G and GS are given in Mitchell and Johns (1997).

Observational near-surface temperature trends are calculated from an updated dataset of blended land surface air temperatures and SSTs (Parker et al. 1994). Monthly average near-surface temperature fields are converted to annual mean fields with the requirement that there must be at least eight months with data present. A check was made to ensure that this criterion did not lead to any significant bias toward a particular season; annual mean values were also calculated with the requirement that all four seasonal means should be present (computed from at least one month). Our results were found to be insensitive to the criteria used. Trends are calculated from the annual mean fields with the requirement that at least half of the data must be present at each grid point. The pattern of observed 50-yr trends calculated in this way up to 1995 and the corresponding missing data mask are shown in Fig. 1. Once the trends have been calculated, missing data in the observations are replaced by the (area weighted) global mean value. Kim and North (1993) also projected modeled and observed fields (in their case temperature anomalies) onto spherical harmonics, and they also chose to substitute the global mean for missing values in the observations. We also tried another method in which missing data are filled in by interpolation; both methods gave similar detection results.

The model data are masked in the same way as the observational data. For each annual mean model field, at each point where there is no observational data for that year (calculated as described above), the model data is replaced by a missing data indicator. Trends are then calculated in the same way as for the observational data and missing trend data are then replaced by their global mean value.

Uncertainty analysis is performed using 1000 yr of data from CTL. The ability of the control simulation to represent the natural internal variability of the ocean–atmosphere system is discussed in detail by Tett et al. (1997). Temperature trends over N yr from CTL are calculated beginning from the first yr of the control run and starting at intervals of 10 yr thereafter until the end of the 1000-yr run. For N greater than 10 yr, therefore, the trends are overlapping.

The original field of near-surface temperature trends, ΔT, is represented by a truncated series of spherical harmonics:

$$\Delta T(N, \lambda, \phi) = \sum_{l=0}^{M} \sum_{m=-1}^{l} \Delta T_{m,l}(N) P_{m,l}(\mu) e^{im\lambda}, \quad (1)$$

Unauthenticated | Downloaded 06/13/21 11:50 PM UTC

where m is the zonal wavenumber and l is the total wavenumber. A triangular truncation is used; $\mu = \sin\phi$, ϕ is latitude, λ is longitude, N is the number of years over which tendencies are calculated, and $P_{m,l}(\mu)$ are the Legendre polynomials. The spherical harmonic coefficients $\Delta T_{m,l}$ are calculated according to a method due to Machenhauer and Daley (1972).

The spherical harmonics have the property that all harmonics with the same value of total wavenumber, 1, have the same spatial scale. The fingerprint at a particular spatial and temporal scale, $\mathbf{f}(N, l)$, is the vector of spherical harmonics:

$$\mathbf{f}(N, l) = [\Delta T^M_{l,l}(N), \Delta T^M_{l-1,l}(N), \ldots, \Delta T^M_{-l,l}(N)]^{\mathrm{T}}, \quad (2)$$

where, for a triangular truncation, $\mathbf{f}(N, l)$ contains $2l + 1$ complex spherical harmonic coefficients for N-yr trends from either the G or GS simulations. The spherical harmonic coefficients have the symmetry property that the l coefficients having $m > 0$ are the complex conjugates of the l coefficients with $m < 0$.

a. Current model skill

Near-surface temperature trends up to 1995 from the model and from the observations are used to determine the spatial and temporal scales on which HADCM2 currently has skill at modeling observed patterns of near-surface temperature change.

1) DETECTION

The current skill of the model is assessed by comparing temperature trends over N yr to 1995 calculated from the G and GS ensemble means with the observed N-yr trends also to 1995. We consider 10-, 30-, and 50-yr trends. The observational data mask is applied to the model data; for each year to 1995, model values at grid points that have missing data in the observational record are replaced by the missing data indicator and trends are calculated in the same way as the observed trends with missing trend data replaced by the global mean values. The observational data mask is also applied to trends taken from the control simulation.

The detection variable $d(N, l)$ is formed by applying the fingerprint to the observations by calculating the dot product:

$$d(N, l) = \mathbf{f}(N, l) \cdot \mathbf{O}(N, l) = \sum_{m=-l}^{l} f_m(N) O_m(N), \quad (3)$$

where $\mathbf{f}(N, l)$ is calculated from the N-yr G or GS ensemble mean trends to 1995, $\mathbf{O}(N, l)$ is the corresponding vector of observed spherical harmonic coefficients and $f_m(N)$, $O_m(N)$ are elements of $\mathbf{f}(N, l)$, and $\mathbf{O}(N, l)$ given by $\Delta T_{m,l}(N)$ where the N-yr temperature trends are calculated from the model and from the observations, respectively.

We test the null hypothesis that the observed climate change originates from natural variability by calculating the distribution of

$$\tilde{d}(N, l) = \mathbf{f}(N, l) \cdot \tilde{\mathbf{V}}(N, l), \quad (4)$$

where $\tilde{\mathbf{V}}(N, l)$ is the distribution of vectors of spherical harmonic coefficients expected from natural variability. This distribution is made up of N-yr trends taken from the 1000-yr control run computed as described above. Therefore the nth member of the distribution $\tilde{d}(N, l)$ is given by

$$d_n(N, l) = \mathbf{f}(N, l) \cdot \mathbf{V}_n(N, l) \quad (5)$$

where $\mathbf{V}_n(N, l)$ is taken from the nth segment of the control run. Our detection results are therefore dependent on the model's ability to simulate realistically unforced climate variability.

The detection variable is the dot product of two vectors of length $2l + 1$. To avoid a growth in the value of the detection variable with increasing l, the fingerprint pattern $\mathbf{f}(N, l)$ is weighted by $1[(2l + 1)]^{-1/2}$. The weighted dot product of two vectors whose elements have equal variance then has variance that is independent of l.

We now determine whether $d(N, l)$ lies within the distribution $\tilde{d}(N, l)$. We choose a risk $1 - p$ of a type-1 error (that the null hypothesis is true but is rejected) and calculate the smallest interval Ω for which the probability that $\tilde{d}$ lies within that interval is given by p. The null hypothesis is then rejected if the detection variable for the observed climate change $d(N, l)$ is contained outside the interval Ω. The detection variable is expected to be positive if greenhouse warming is occurring and therefore a one-tailed test is performed. Here we choose $p = 0.95$. We claim detection if $d(N, l)$ is greater than the 95 percentile of $\tilde{d}(N, l)$.

The same analysis is carried out over a band of total wavenumbers as well as for each individual wavenumber. The results then yield detection results for a range of spatial scales for each particular trend length, N.

2) CONSISTENCY

A test is made to determine whether near-surface temperature trends from the G and GS simulations are inconsistent with the observed trends over a range of spatial scales. We calculate the distribution of fingerprints calculated from the mean of three members of the four-member ensemble applied to the observations. This distribution, $\tilde{d}_O$, contains four members:

$$\tilde{d}_O = \{\mathbf{f}_{\overline{1,2,3}}(N,L) \cdot \mathbf{O}(N,L),\ \mathbf{f}_{\overline{1,2,4}}(N,L) \cdot \mathbf{O}(N,L),\ \mathbf{f}_{\overline{1,3,4}}(N,L) \cdot \mathbf{O}(N,L),\ \mathbf{f}_{\overline{2,3,4}}(N,L) \cdot \mathbf{O}(N,L)\}, \quad (6)$$

where the subscripts 1, 2, 3, and 4 identify the four members of the G or GS ensembles and $\mathbf{f}_{\overline{1,2,3}}$, for example, is the fingerprint vector calculated from the mean of the first three members of the four-member ensemble.

Unauthenticated | Downloaded 06/13/21 11:50 PM UTC

The notation (N, L) indicates that the calculation is made over a band of total wavenumbers appropriate to a range of spatial scales for a particular trend length, N. This distribution is compared with the distribution in which the observations are substituted by the remaining ensemble member not included in the three-member mean. This distribution, $\tilde{d}_M$, contains four members:

$$\tilde{d}_M = \{\mathbf{f}_{\overline{1,2,3}}(N,L)\cdot\mathbf{f}_4(N,L),\ \mathbf{f}_{\overline{1,2,4}}(N,L)\cdot\mathbf{f}_3(N,L),\ \mathbf{f}_{\overline{1,3,4}}(N,L)\cdot\mathbf{f}_2(N,L),\ \mathbf{f}_{\overline{2,3,4}}(N,L)\cdot\mathbf{f}_1(N,L)\}, \quad (7)$$

where $\mathbf{f}_1(N, L)$, for example, is the fingerprint vector calculated from the first member of the four-member ensemble. A Kolmogorov–Smirnov test (Press et al. 1992) is made to determine whether the two datasets $\tilde{d}_O$ and $\tilde{d}_M$, each of which has four members, are drawn from the same distribution function. Since this is not a very powerful test, a Student's t-test has also been carried out to determine whether there is a significant difference between the means of the two datasets.

b. Perfect model study

In the future, as the signal of climate change becomes stronger relative to the noise of natural internal variability, the range of spatial and temporal scales on which we are able to detect climate change is expected to increase. The aim of this section is to determine on what scales future climate change would be detectable in principle if we were able to develop a model that simulates the observed climate change perfectly to within the limits of natural variability. The strategy we adopt is to calculate a probability of future detection, an approach that has been taken by Allen et al. (1994a) and Allen et al. (1994b) in their investigation of the prospects for climate change detection using satellite SST observations.

Here we calculate all possible permutations in which one member of each ensemble (G or GS) is taken to represent the observations and another member of the same ensemble is taken to represent a model realization of the future temperature trend. It is possible to calculate four fingerprint patterns $\mathbf{f}_i(N, l)$ (where i runs from 1 to 4) from the four ensemble members and these are applied to the other ensemble members that represent three possible realizations of the observational N-yr trends at a particular spatial scale (corresponding to total wavenumber, l). For each of the four fingerprint vectors, therefore, three detection variables are calculated from

$$d_{ij}(N, l) = \mathbf{f}_i(N, l)\cdot\mathbf{f}_j(N, l), \quad (8)$$

where $j \neq i$. The three detection variables for each value of i are compared with the distribution expected from natural variability:

$$\tilde{d}_i(N, l) = \mathbf{f}_i(N, l)\cdot\tilde{\mathbf{V}}(N, l), \quad (9)$$

where $\tilde{\mathbf{V}}(N, l)$ is the distribution of vectors of spherical harmonic coefficients that are expected from natural variability and that are taken from the 1000-yr control run as before. In other words, the kth member of the distribution $\tilde{d}_i(N, l)$ is given by

$$d_{i,k}(N, l) = \mathbf{f}_i(N, l)\cdot\mathbf{V}_k(N, l), \quad (10)$$

where $\mathbf{V}_k(N, l)$ is taken from the kth segment of the control run. The total number of occurrences when $d_{ij}(N, l)$ exceeds the 95 percentile of the distribution $\tilde{d}_i(N, l)$ (summed over the four values of i and the three values of j for each i) divided by 12 gives the probability of detection for each N, l.

Detection may require more than one total wavenumber, l, if the signal projects onto a range of spherical harmonics with different values of l. The same analysis is therefore carried out for a sequence of wave bands representing a range of spatial scales.

c. Variability

The results presented here, in common with other detection work, assume that the model adequately simulates internal climate variability. In this section we assess the difference between observed and modeled variability by comparing the power in the observations with the power in the control run for 50-yr trends over a range of spatial scales. However, part of the power in the observations results from the signal of climate change; we therefore remove the power associated with this signal by calculating the observed power in directions orthogonal to the signal. The component of $\mathbf{O}(N, l)$ in directions orthogonal to the signal $\mathbf{S}(N, L)$ is given by

$$\mathbf{O}(N, L) - \frac{[\mathbf{O}(N, L)\cdot\mathbf{S}(N, L)]\mathbf{S}(N, L)}{\mathbf{S}(N, L)\cdot\mathbf{S}(N, L)}, \quad (11)$$

and the power is given by the square of this expression:

$$\mathbf{O}(N, L)\cdot\mathbf{O}(N, L) - \frac{[\mathbf{O}(N, L)\cdot\mathbf{S}(N, L)]^2}{\mathbf{S}(N, L)\cdot\mathbf{S}(N, L)}. \quad (12)$$

The signal $\mathbf{S}(N, l)$ is chosen from a linear combination of members of the G and GS ensembles that maximizes the projection of the observations onto the signal. This is achieved by computing the multivariate regression between the observations $\mathbf{O}$ and the members of the G and GS ensembles $\mathbf{S}_i$:

$$\mathbf{O} = \sum_i a_i\mathbf{S}_i + \epsilon, \quad (13)$$

where $\mathbf{O}$ and $\mathbf{S}_i$ are vectors that include all the spherical harmonic coefficients appropriate to the temporal scale and range of spatial scales being considered, ϵ is the residual, and the index i is over the ensemble members. The solutions a_i minimize ϵ, so that the best estimate of the signal present in the observations is

$$\mathbf{S}(N, L) = \sum_i a_i\mathbf{S}_i. \quad (14)$$

The power in the observations orthogonal to the sig-

Unauthenticated | Downloaded 06/13/21 11:50 PM UTC

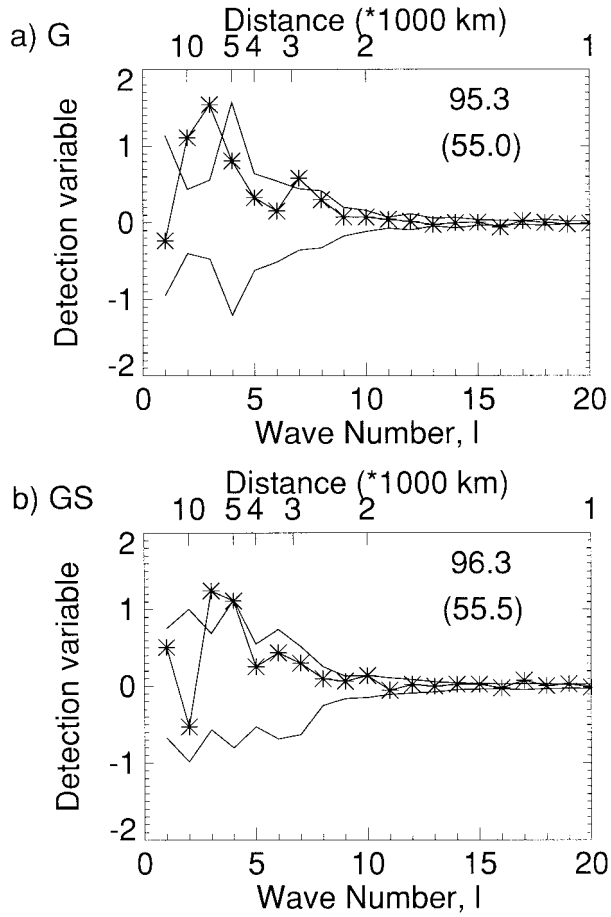

FIG. 2. The d (N, l) (line with stars) and 95% and 5% confidence intervals for $\tilde{d}(N, l)$ where $f(N, l)$ is derived from (a) the G ensemble mean, and (b) the GS ensemble mean. Here, $d(N, 0)$, that is, the global mean, is shown in the panel with the 95% confidence interval for $\tilde{d}(N, 0)$ shown in brackets. The detection variable $d(N, l)$ at each wavenumber, l, is expressed as a percentage of the detection variable calculated for all wavenumbers including the global mean.

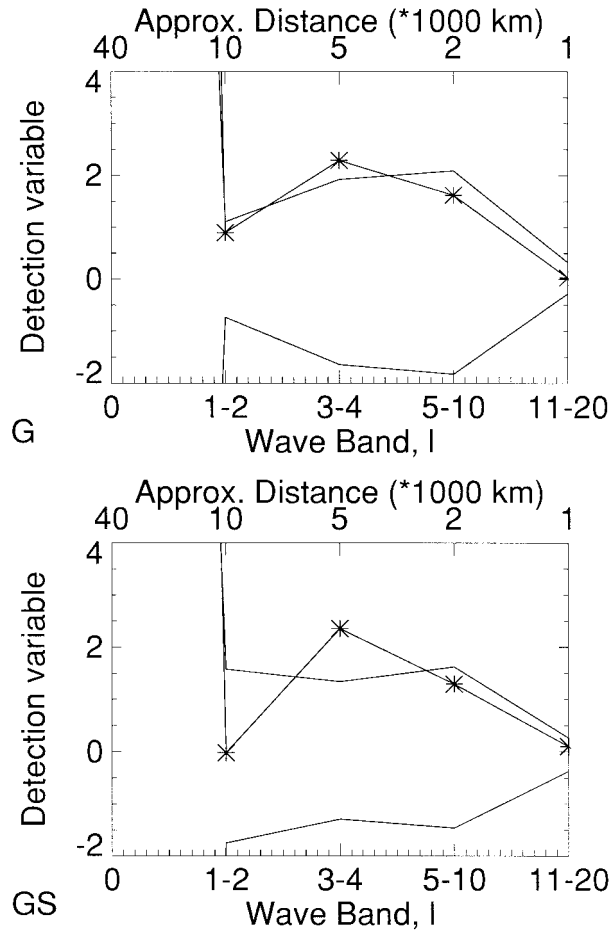

FIG. 3. As in Fig. 2 except that now d and $\tilde{d}$ are calculated over the wave band that includes the total wavenumbers as described on the bottom axis.

nal is compared with the power in the control, also in directions orthogonal to the signal. The relevant distribution calculated from the control run has members given by

$$\mathbf{V}(N, L) \cdot \mathbf{V}(N, L) - \frac{[\mathbf{V}(N, L) \cdot \mathbf{S}(N, L)]^2}{\mathbf{S}(N, L) \cdot \mathbf{S}(N, L)}, \quad (15)$$

where $\mathbf{V}(N, L)$ is a vector of spherical harmonic coefficients representing N-yr trends taken from the control run as described above.

3. Results

The aims that we set out in the introduction, namely to identify those scales on which HADCM2 currently has skill, on which climate change is detectable in future and on which the model adequately simulates the climate's internal variability, are addressed in this section.

a. On what scales does the model currently have skill?

Here we identify those scales on which HADCM2 currently has skill at modeling observed patterns of near-surface temperature change.

1) DETECTION

The methodology we use to determine the spatiotemporal scales on which the level of agreement between observed and modeled patterns is significantly different from zero and hence on which we claim climate change detection has been described in section 2a. The detection variable formed by applying the fingerprint to the observations is compared with the distribution expected from natural variability; this is shown in Fig. 2 when calculated for each total wavenumber and in Fig. 3 when calculated over a sequence of wave bands to represent a range of scales. For both G and GS, the global mean accounts for over 95% of the detection variable calculated over all scales including the global mean (Fig.

Unauthenticated | Downloaded 06/13/21 11:50 PM UTC

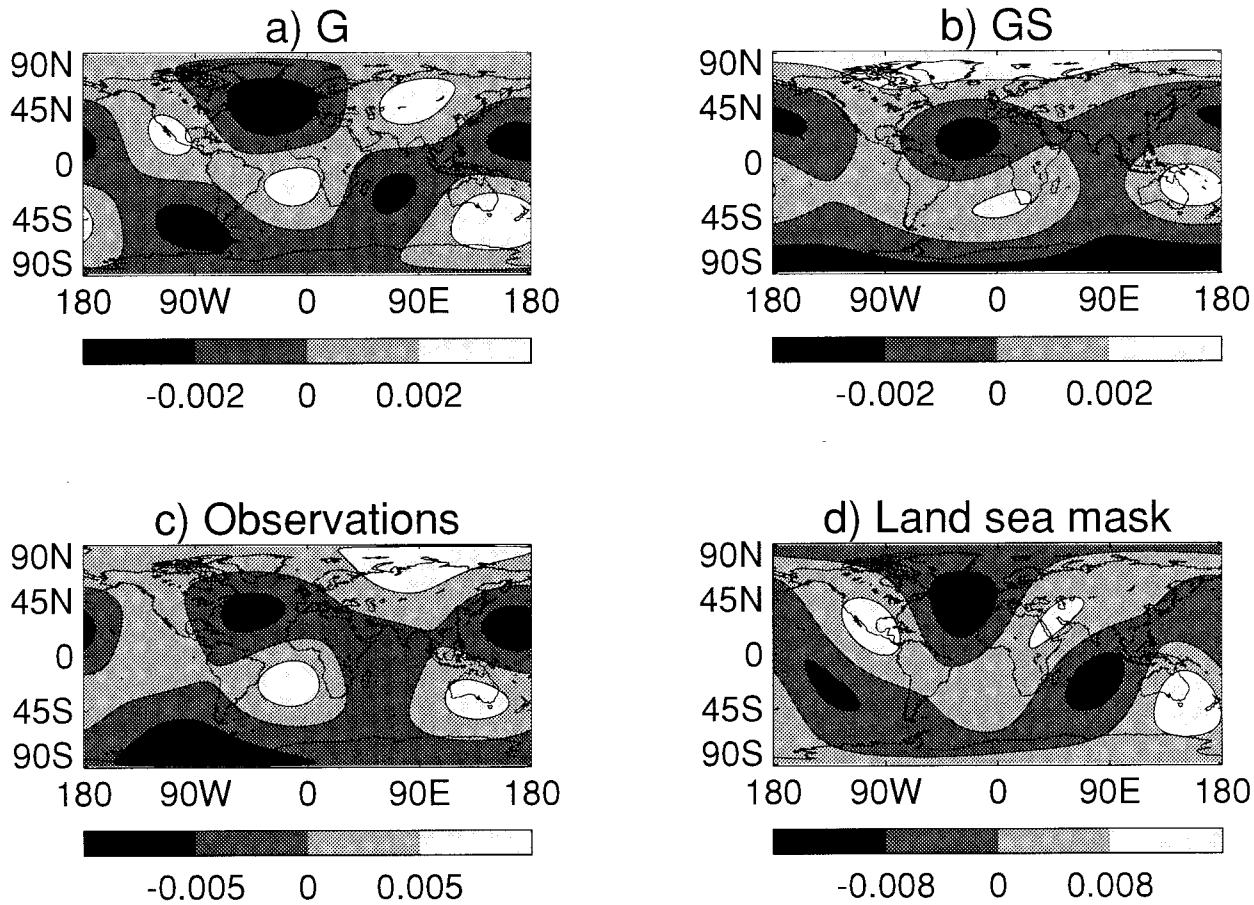

FIG. 4. Gridpoint fields reconstructed from the spherical harmonic coefficients with total wavenumber 3 for (a) G, (b) GS, and (c) observational 50-yr trends to 1995 and (d) the land–sea mask.

2). For global mean trends, the detection variables derived from both G and GS are statistically significant at the 5% level. At subglobal scales there are statistically significant values of the detection variable for large spatial scales (small total wavenumber). The detection variable $d(N, l)$ is significant at the 5% level for total wavenumbers 2 and 3 for the G fingerprint (Fig. 2a) and for total wavenumber 3 for the GS fingerprint (Fig. 2b). These results indicate that little is gained in including wavenumbers greater than total wavenumber 4, corresponding to spatial scales smaller than 5000 km, when making a test for detection with current near-surface temperature trends. When calculated over four wave bands representing subglobal scales, the detection variable is only significant for one wave band corresponding to an approximate length scale of 5000–10 000 km (Fig. 3).

The most significant departures of the detection variable $d(N, l)$ from the distribution of $\tilde{d}$ are at $l = 3$ for both GS and G. Gridpoint fields of 50-yr near-surface temperature trends were reconstructed from the spherical harmonics with $l = 3$ from G and GS and compared with the equivalent observational field (Figs. 4a,b,c, respectively). The wave 3 fields for G, GS, and the observations all have cold regions in the North Atlantic and the North Pacific, and warm regions in central Asia–Siberia, the South Atlantic, and Australia. Due to the greater heat capacity of the oceans, land areas are expected to respond faster to climate forcing than sea areas. To see whether there is a fingerprint of this differential climate response, we projected the land–sea mask onto spherical harmonics. A comparison with the wave 3 field for the land–sea mask (Fig. 4d) indicates that the wave 3 pattern seen in G (Fig. 4a) reflects to a large extent the land–sea contrast. Indeed when fingerprint vectors are calculated from projections of the land–sea mask onto spherical harmonics, the most significant value of the detection variable is at wave 3 (not shown). The wave 3 pattern for GS (Fig. 4b) shows a combination of the land–sea contrast and a band of cooling in Northern Hemisphere midlatitudes.

The results for 10- and 30-yr trends are summarized in Table 1, which shows values of the detection variable at the global mean and summed over the first 20 wavenumbers. The values for 50-yr trends, which were discussed above and that are shown graphically in Figs. 2

Unauthenticated | Downloaded 06/13/21 11:50 PM UTC

TABLE 1. Detection variable and 95 percentile (in parentheses) of distribution expected from natural variability. Detection variables that are statistically significant at the 5% level are shown in bold.

	N = 10 yr	N = 30 yr	N = 50 yr
G (1 = 0)	79.4 (158.9)	**100.4** (51.8)	**95.3** (55.0)
GS (1 = 0)	75.3 (150.7)	**95.8** (49.5)	**96.3** (55.5)
G (1 = 1–20)	20.6 (22.2)	−0.4 (4.5)	**4.7** (3.7)
GS (1 = 1–20)	24.7 (36.7)	4.2 (5.7)	**3.7** (2.8)

TABLE 2. Are the modeled and observed 50-yr trends inconsistent? Here, Y (N) indicates that it is possible (is not possible) to reject the null hypothesis (with a risk of 5%) that the two sets of data are drawn from the same distribution according to a Kolmogorov–Smirnov test.

Wave no.	Approx distance (km)	G	GS
0	40 000	Y	N
1–4	20 000–5 000	N	N
5–10	5 000–2 000	N	N
11–20	2 000–1 000	Y	N

and 3, are also shown for comparison. At the global mean, the detection variables derived from 30- and 50-yr trends are statistically significant at the 5% level for both G and GS. Ten-year global mean trends are not significant. This agrees with Santer et al. (1995), whose C(t) statistic provides information on global mean temperature trends and that shows significance for trend lengths of 20 yr and greater. At subglobal scales, as already seen, the detection variables for 50-yr trends from both G and GS are significant; however, the detection variables for 10- and 30-yr trends are not significant at the 5% level at subglobal scales. In summary, statistically significant values of the detection variable (i.e., projections of the signals onto the observations that are outside the range expected from natural variability) are seen only at large spatial scales (length scales greater than 5000 km) and for long trend lengths (multidecadal trends).

2) CONSISTENCY

Are 50-yr trends from the G and GS ensemble mean simulations consistent with the observed 50-yr near-surface temperature trend to 1995? Table 2 shows the results of a Kolmogorov–Smirnov test, at a significance level of 5%, to determine whether the two datasets $\tilde{d}_O$ and $\tilde{d}_M$ are drawn from the same distribution function (see section 2a). The null hypothesis that the two sets of data are drawn from the same distribution cannot be rejected at any spatial scale for GS. This is not the case for G at the global mean or at spatial scales smaller than 2000 km.

We also carried out a Student's t-test (at the same 5% significance level) to see on what scales there are significant differences in the means of the two distributions. [The Student's t-test in its standard form only applies if the two distributions have the same variance. An F test showed that the two distributions did not have significantly different variances for all wave bands except for global mean G. For this wave band an unequal-variance t test was applied (Press et al. 1992), and it gave the same results as the standard t test.] The Student's t-test gave the same results as the Kolmogorov–Smirnov test for all wave bands except global mean GS, where the Student's t-test indicated that there was a significant difference between the global mean 50-yr trend from the GS simulation and from the observations.

The G simulation significantly overestimates the observed increase in global mean temperatures over the 50 yr to 1995 (0.74 K for the G ensemble mean and 0.31 K for the observations). The 50-yr trend to 1995 from GS is closer to the observed trend but is also overestimated (0.48 K for the GS ensemble mean); this difference is significant according to the Student's t-test but is not significant according to the less powerful Kolmogorov–Smirnov test. Mitchell et al. (1995) showed that including the cooling due to sulfate aerosols gives a simulation closer to the observations in recent decades than a simulation that includes only the effects of increasing greenhouse gases. The overestimate of the global mean 50-yr trend to 1995 by GS is related to GS being too cool in the 1940s and 1950s (see Mitchell et al. 1995). There are large uncertainties in the sulfate aerosol forcing in the model, and the simulations do not include natural factors such as changes in solar output and volcanic eruptions of dust into the stratosphere that may help to explain the discrepancy between modeled and observed global mean temperatures during the 1940s and 1950s.

The small-scale structure (less than 2000 km) of G is also inconsistent with the observations. Calculating fingerprint vectors from the land–sea mask for each spatial scale and comparing its projection onto the G and GS fingerprint vectors (not shown) indicates that G contains a greater projection of the land–sea mask related signal onto small scales than GS. In the next section we show that the emergence of the pattern of land–sea contrast in future near-surface 50-yr temperature trends leads eventually to detection at all scales. As we have seen, current 50-yr trends to 1995 do not show detection at small scales; this is consistent with the fact that the land–sea pattern does not emerge clearly in the observed 50-yr temperature trend to 1995 (compare Fig. 8c and Fig. 8d), and there is no significant projection of the land–sea mask onto the observations at small scales. It seems likely therefore that the small-scale structure of the 50-yr G trends has significantly too strong a signal from the land–sea contrast.

b. On what scales is climate change detectable?

The future probabilities of detection for the GS and G simulations for 50-yr trends at 10-yr intervals to 2095 are shown in Fig. 5 when expressed as a function of total wavenumber (l) and in Fig. 6 when expressed over

Unauthenticated | Downloaded 06/13/21 11:50 PM UTC

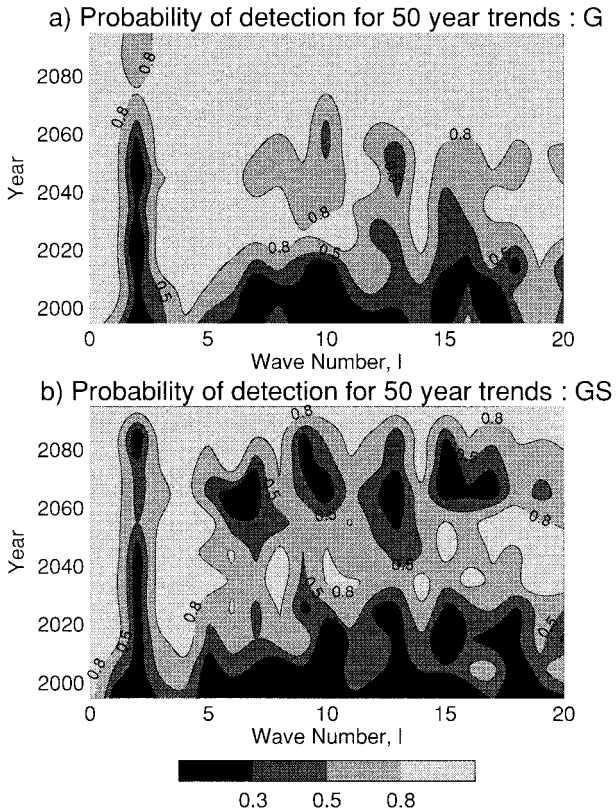

FIG. 5. Probability of detection for 50-yr trends at 10-yr intervals to 2095 for (a) G and (b) GS, calculated for each total wavenumber 1. Contours are at 0.3, 0.5, and 0.8.

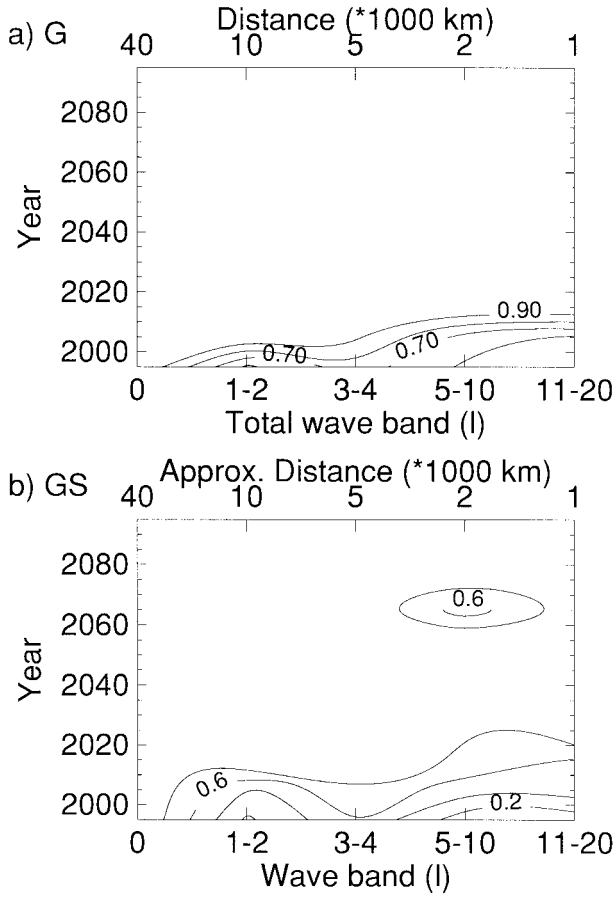

FIG. 6. Probability of detection for 50-yr trends at 10-yr intervals to 2095 for (a) G (contour interval 0.1) and (b) GS (contour interval 0.2), expressed as a function of wave band.

the same four subglobal scale wave bands as Fig. 3. Figure 5 is a noisier picture than Fig. 6, and values are generally higher when probabilities are calculated over wave bands than individual waves. Many of the detection variables for future trends calculated for individual wavenumbers corresponding to small spatial scales exceed the 70 or 80 percentiles but not the 95 percentiles of the distribution expected from natural variability. When grouped into bands however, a high proportion of the corresponding detection variables do exceed the 95 percentile, since the detection variable for many of the wavenumbers making up the wave band are in the upper quartile of the distribution expected from natural variability.

Detection is very probable for global mean 50-yr trends at all times for both G and GS. At subglobal scales, detection probabilities for G (Fig. 5a) are initially largest for wavenumbers 3 and 4 and then for a gradually increasing range of wavenumbers during the first half of the next century. By the latter half of the century detection probabilities exceed 0.8 for the great majority of wavenumbers at all spatial scales. Detection probabilities at subglobal scales for GS (Fig. 5b) are highest predominately at wavenumbers 1, 3, and 4. High probabilities of detection for the majority of wavenumbers at all spatial scales only emerge for the 50-yr trend to 2095. At small scales, detection probabilities generally increase in the early part of the next century before decreasing to reach a minimum and increasing again toward the end of the century.

This pattern of increasing, decreasing, and increasing probabilities at subglobal scales can be understood as the transient response of the land–sea contrast near-surface temperature trend to a changing forcing. The transient nature of the response of a simple coupled climate system has been discussed by Kim et al. (1992). To illustrate this, following an approach similar to Kim et al. (1992), we consider a simple climate model:

$$\frac{dT}{dt} = \frac{1}{C}(F - T), \tag{16}$$

where C is the lag for ocean or land areas, F is the final temperature response for a particular forcing, and T is the transient temperature response. Taking the forcing histories used in the model simulations (Mitchell and Johns 1997) and a climate sensitivity for the model of 2.8 K for a doubling of CO_2 (forcing of 4.1 W m^{-2}) (C. Senior 1998, personal communication) we solve (16) for land and ocean areas. For land areas, we multiply

Unauthenticated | Downloaded 06/13/21 11:50 PM UTC

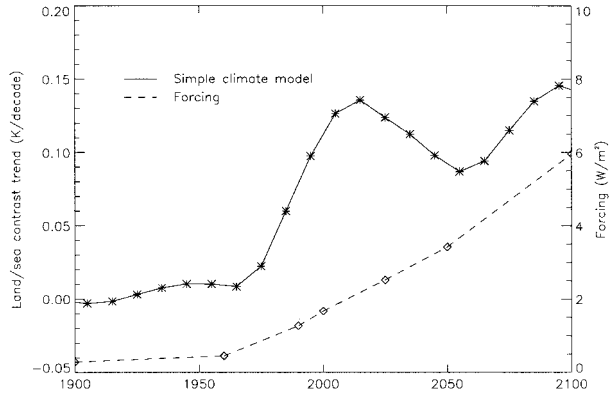

FIG. 7. Trend in surface temperature land–sea contrast calculated using a simple climate model.

the final temperature response F by a factor that takes account of the greater climate sensitivity of the land relative to the ocean as a result of differences in cloud feedback, evaporative cooling potential, and changes in surface albedo (Murphy 1995). We obtain two time series for land and ocean areas respectively from which we calculate the 50-yr trends in land–sea temperature contrast predicted by this simple model. The results are shown in Fig. 7 where we take values for the parameters in the simple model of 20 yr for the ocean lag (Kim et al. 1992), 1 yr for the land lag, and 1.5 yr for the land sensitivity (Murphy 1995). The simple climate model of (16) illustrates how the magnitude of the land–sea contrast trend changes with time as a result of the changing forcing due to greenhouse gases and sulfate aerosols. Figure 7 shows the land–sea contrast trend increasing through the latter part of the twentieth and early part of the twenty-first century before decreasing in the middle of the next century and increasing once more in the latter half of that century. This double-peak structure in land–sea contrast temperature trends is also seen in the probability of detection for future 50-yr temperature trends shown in Fig. 5b. The land–sea contrast temperature trends and hence the probability of detection at subglobal scales are responding to a changing forcing due to greenhouse gases and sulfate aerosols that increases sharply in 1960 (increase in rate of change of forcing due to both CO_2 and sulfate aerosols) and 2050 (sulfate aerosol forcing flattens off) (see Fig. 7 and Fig. 2a of Mitchell and Johns 1997).

For 50-yr trends to 1995, detection probabilities at subglobal scales are largest for the wave band with wavenumbers 3 and 4 (>0.7 for G, >0.5 for GS), and smallest for the small-scale waves with $l = 11 - 20$ (<0.6 for G, <0.2 for GS) (Fig. 6). The results for GS are consistent with the current detection variables calculated using observed 50-yr trends shown in Fig. 3b; detection is observed only where the detection probabilities calculated from the perfect model experiment are greater than 0.5. For G, the perfect model results indicate that detection is more likely than not to have been observed in G at scales smaller than 5000 km, although the detection variable calculated using observed 50-yr trends is not significant. However, we showed in section 3a that the small-scale structure of the 50-yr G trend to 1995 is inconsistent with the observations.

For future trends, the probabilities of detection increase with time and by the middle of the next century, detection over all four subglobal scale wave bands is extremely likely. Thus, if the natural climate system behaves like HADCM2 and HADCM2 is able to prefectly capture the signal of future climate change, detection is highly probable by the middle of the next century at all spatial scales larger than 1000 km. The minimum in probabilities of detection for GS seen after 2060 (Fig. 6b) corresponds to the minima seen for individual waves in Fig. 5b at about this time. As was shown above, this double peak structure is a result of the transient nature of the climate response to a changing forcing.

For the rest of this section we concentrate on GS. The current lack of detection at small spatial scales is illustrated by reconstructing gridpoint fields from all the small-scale waves with wavenumbers between 10 and 20 and focusing on a particular geographical region for clarity. Figure 8 shows the results for the region around Australia; the same plots for other regions, such as Africa or Europe, for example, tell the same story. A comparison between Fig. 8a and Fig. 8d shows that there is little correlation between observed and modeled trends to 1995. On the other hand, the much higher degree of pattern correlation between modeled trends to 2050 and the land–sea mask (Fig. 8b and Fig. 8c) illustrates the increasing manifestation of the land–sea contrast in the small-scale structure of the modeled temperature response. This signal is seen in all ensemble members and is responsible for the highly significant agreement between ensemble members that emerges at small spatial scales.

Probabilities of detection for 10-yr trends remain low (<0.2) at all scales and at all times except for global mean trends in the latter half of the twenty-first century, the probability of detection for which averages 0.5. The probability of detection for 30-yr global mean trends also remains low at all times (<0.5) except for the global mean (which is always 1.0), for the largest two waves ($l = 1, 2$) after 2075 (>0.8), and for all four wave bands after 2095 (>0.7).

c. Variability

The ability of HADCM2 to simulate observed climate variability has been assessed at a range of spatial scales. The power of the observed 50-yr near-surface temperature trends is compared with the distribution of power of 50-yr trends from the control run in directions orthogonal to the climate change signal. The signal is estimated from the G and GS simulations as described

Unauthenticated | Downloaded 06/13/21 11:50 PM UTC

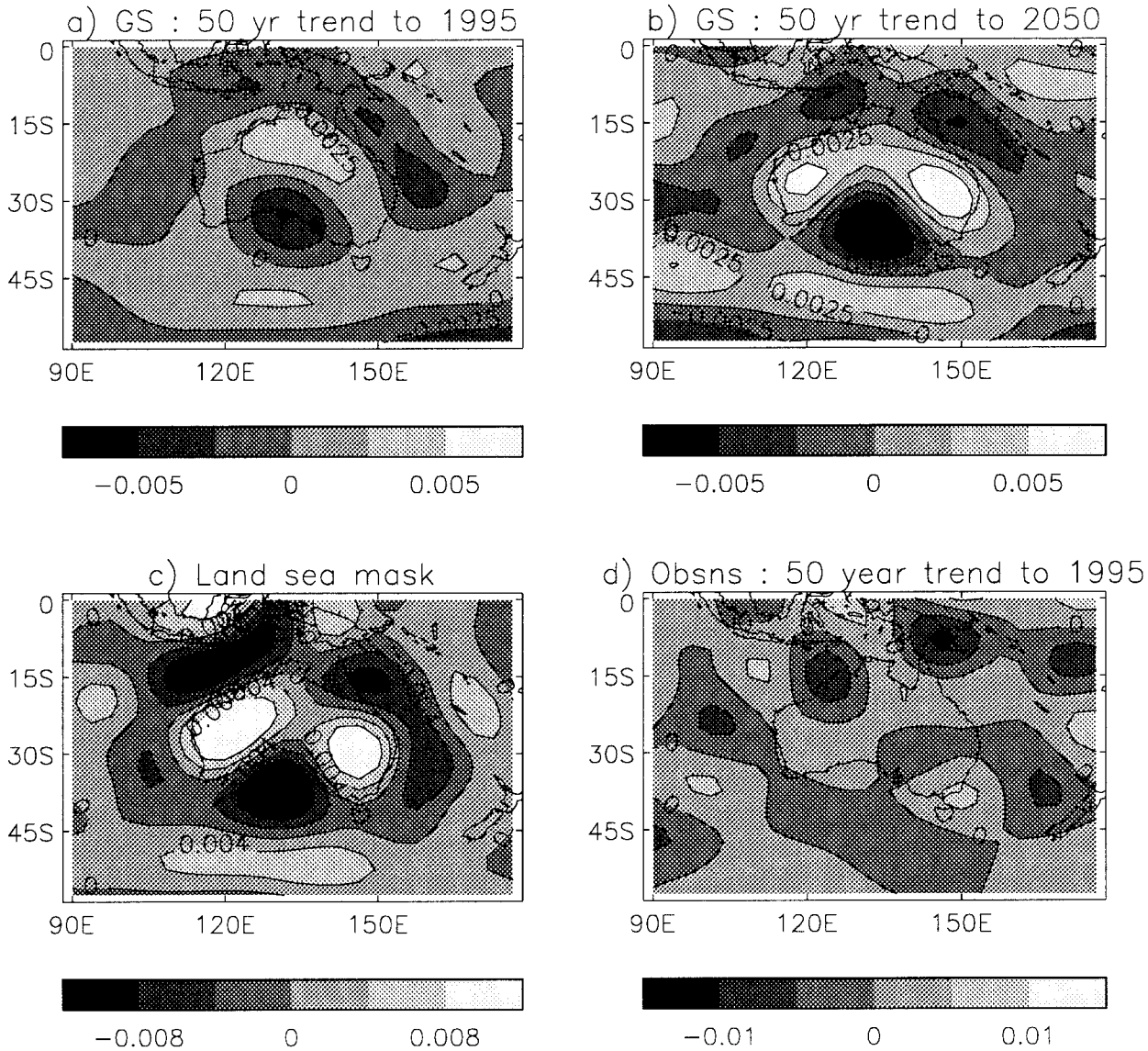

FIG. 8. Gridpoint fields reconstructed from spherical harmonic coefficients with total wavenumbers 10–20 for (a) GS 50-yr trends to 1995, (b) GS 50-yr trends to 2050, (c) land–sea mask, and (d) observed 50-yr trends to 1995.

TABLE 3. Is the variability in the observations (measured by the power orthogonal to the signal estimated from G and GS) greater than in the control simulation? Here, Y indicates that the observed power is greater than the 95 percentile of the distribution of power calculated from the control run, that is, that observed variability is significantly greater than the modeled variability at that scale. Here, N indicates that observed variability is not significantly greater than the modeled variability.

Wave no.	Approx distance (km)	Observed variability > control variability?
1–4	20 000–5 000	N
5–10	5 000–2 000	N
11–20	2 000–1 000	Y

in section 2c. If the observed power at a particular scale is greater than the 95 percentile of the power distribution calculated from the control run, the observed variability can be said to be significantly greater than the modeled variability at that scale. For the two wave bands corresponding to spatial scales greater than 2000 km (Table 3), there is no significant difference between modeled and observed variability. For the wave band corresponding to scales less than 2000 km; however, the model significantly underestimates the observed variability.

In climate change detection and attribution studies, it is crucial to question whether the model simulation of internal climate variability is adequate to quantify uncertainty. The results reported here indicate that models do not adequately represent variability at small spatial

Unauthenticated | Downloaded 06/13/21 11:50 PM UTC

scales. This is particularly relevant to optimal fingerprinting techniques (e.g., Hegerl et al. 1996) that rotate the fingerprint of climate change into directions with small noise, therefore enhancing the signal to noise ratio. If the optimal fingerprint is rotated into directions in which variability is poorly represented by the model, detection may be erroneously claimed. A simple consistency check to ensure that the problem of too little variance at small spatial scales does not compromise detection results from optimal fingerprinting methods has been proposed by Allen and Tett (1998).

4. Conclusions

A study has been made of the spatial and temporal limits to the detection of climate change by projecting near-surface *N*-yr temperature trends from the G and GS simulations onto spherical harmonics. The results show that climate change can be detected on the global mean scale for 30- and 50-yr trends but not for 10-yr trends. At subglobal scales, climate change can be detected for 50-yr trends and only for large spatial scales. For 50-yr trends the results indicate that there is no improvement to the detection ability of the forced simulations beyond total wavenumber 4 or spatial scales greater than 5000 km. The largest subglobal scale signal for both G and GS is at total wavenumber 3. This signal is associated largely with the land–sea contrast; when the land–sea mask is used to derive a fingerprint pattern, the largest signal is also at wave 3.

A consistency test, which compares patterns of observed and modeled temperature trends, shows that GS is not inconsistent with the observations at all spatial scales. Here, G, on the other hand, is inconsistent with the observations at spatial scales smaller than 2000 km. This appears to result from a greater projection of the land–sea contrast onto small scales in G than is seen in the observations.

A study, in which one member of the four-member ensemble of simulations is taken to represent a ''perfect'' model and the other three members are taken to represent the observations, shows that detection for 50-yr trends becomes highly probable at all spatial scales by the middle of the next century. For the next 20 yr, detection is most likely at large spatial scales. On the other hand, detection is very unlikely at all subglobal scales for 10-yr trends; the same is true for 30-yr trends until 2075 when the probability of detection for the largest two waves (spatial scales larger than 10 000 km) becomes high.

In this work, we have considered simulations of HADCM2 that include forcing due to greenhouse gases and sulfate aerosols. We have neglected a number of other forcing factors including naturally occurring changes in solar output and volcanic dust. Volcanoes have a significant impact on the climate record as has been demonstrated by North and Stevens (1998) who showed the occurrence in the climate forcing spectrum of strong peaks at frequencies of 0.10 yr^{-1} and 0.11 yr^{-1} due to volcanic eruptions over the last 100 yr. Variations in solar output may also influence climate on timescales of a few decades to a century. Since neither of these natural forcings is included in our analysis, we are not in a position to make claims of attribution of the observed climate change to either anthropogenic or natural causes. In addition, the simulations we consider contain large uncertainties in future sulfate aerosol scenarios and the distribution and radiative properties of sulfate aerosols given an emissions scenario (Mitchell and Johns 1997). Also, the indirect effect of sulfate aerosol has not been considered, nor changes in soot, biogenic aerosols, and ozone.

In common with other studies, these detection results are dependent on the validity of model-based estimates of natural unforced variability. The coupled GCM used in this study underestimates observed unforced variability at spatial scales less than 2000 km. The predictions that detection is very likely at these small spatial scales by the middle of the next century should therefore be treated with caution.

Detection strategies that involve optimization of signal to noise may give high weight to aspects of variability that are unrealistically simulated by the model. The results presented here that show a poor simulation of the high spatial wave number components of variability support the need for a consistency check of the type proposed by Allen and Tett (1998) to be applied to optimal detection strategies.

Acknowledgments. We would like to thank Myles Allen of the Space Science Department at the Rutherford Appleton Laboratory for helpful discussions during the course of this work and Myles Allen, Tim Johns, and John Mitchell for improvements to earlier drafts of this paper. We are grateful to Dr. G. North and an anonymous reviewer whose comments also helped to improve the manuscript. This work was supported by the U.K. Department of the Environment, Transport and the Regions under Contract PECD 7/12/37.

REFERENCES

Allen, M. R., and S. F. B. Tett, 1998: Checking for model consistency in optimal fingerprinting. *Climate Dyn.,* in press.

——, C. T. Mutlow, G. M. C. Blumberg, J. R. Chisty, R. T. McNider, and D. T. Llewellyn-Jones, 1994a: Global change detection. *Nature,* **370,** 24–25.

——, M. J. Panter, G. M. C. Blumberg, C. T. Mutlow, and D. T. Llewellyn-Jones, 1994b: Prospects for global change detection with satellite SST observations. *Proc. Second ERS-1 Symp.,* Hamburg, Germany, Amer. Meteor. Soc., 1103–1108.

Hegerl, G. C., H. V. Storch, K. Hasselman, B. D. Santer, U. Cubasch, and P. D. Jones, 1996: Detecting greenhouse gas-induced climate change with an optimal fingerprint method. *J. Climate,* **9,** 2281–2306.

——, K. Hasselman, U. Cubasch, J. F. B. Mitchell, E. Roeckner, R. Voss, and J. Waskewitz, 1997: Multi-fingerprint detection and attribution analysis of greenhouse gas, greenhouse gas-plus-

Unauthenticated | Downloaded 06/13/21 11:50 PM UTC

aerosol and solar forced climate change. *Climate Dyn.,* **13,** 613–634.

Houghton, J. T., B. A. Callander, and S. K. Varney, 1992: *Climate Change 1992: The Supplementary Report to the IPCC Scientific Assessment.* Cambridge University Press, 200 pp.

Johns, T. C., R. E. Carnell, J. F. Crossley, J. M. Gregory, J. F. B. Mitchell, C. A. Senior, S. F. B. Tett, and R. A. Wood, 1997: The second Hadley Centre coupled ocean–atmosphere GCM: Model description, spinup and validation. *Climate Dyn.,* **13,** 103–134.

Kim, K. Y., and G. R. North, 1993: EOF analysis of surface temperature field in a stochastic climate model. *J. Climate,* **6,** 1681–1690.

——, ——, and J. J. Huang, 1992: On the transient response of a simple coupled climate system. *J. Geophys. Res,* **97,** 10 069–10 081.

Langner, J., and H. Rodhe, 1991: A global three-dimensional model of the tropospheric sulfur cycle. *J. Atmos. Chem.,* **13,** 225–263.

Leung, L.-Y., and G. R. North, 1991: Atmospheric variability on a zonally symmetric land planet. *J. Climate,* **4,** 753–765.

Machenhauer, B., and R. Daley, 1972: A baroclinic primitive equation model with a spectral representation in three dimensions. Inst. Teoretisk Met. Tech. Rep. 4, 63 pp. [Available from Institut for Theoretisk Meteorologi, Haraldsgade 6, DK2200 Copenhagen N, Denmark.]

Mitchell, J. F. B., and T. C. Johns, 1997: On modification of global warming by sulfate aerosols. *J. Climate,* **10,** 245–267.

——, ——, J. M. Gregory, and S. F. B. Tett, 1995: Climate response to increasing levels of greenhouse gases and sulphate aerosols. *Nature,* **376,** 501–504.

Murphy, J. M., 1995: Transient response of the Hadley Centre Coupled Ocean–Atmosphere model to increasing carbon dioxide. Part III: Analysis of global-mean response using simple models. *J. Climate,* **8,** 496–514.

North, G. R., and K.-Y. Kim, 1995: Detection of forced climate signals. Part II: Simulation results. *J. Climate,* **8,** 409–417.

——, and M. J. Stevens, 1998: Detecting climate signals in the surface temperature record. *J. Climate,* **11,** 563–577.

——, K.-J. J. Yip, L.-Y. Leung, and R. M. Chervin, 1992: Forced and free variations of the surface temperature field in a general circulation model. *J. Climate,* **5,** 227–239.

——, K.-Y. Kim, S. S. P. Shen, and J. W. Hardin, 1995: Detection of forced climate signals. Part I: Filter theory. *J. Climate,* **8,** 401–408.

Parker, D. E., P. D. Jones, C. K. Folland, and A. Bevan, 1994: Interdecadal changes of surface temperature since the late nineteenth century. *J. Geophys. Res.,* **99,** 14 373–14 399.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 1992: *Numerical Recipes in Fortran: The Art of Scientific Computing.* 2d ed. Cambridge University Press, 963 pp.

Santer, B. D., K. E. Taylor, T. M. L. Wigley, J. E. Penner, P. D. Jones, and U. Cubasch, 1995: Towards the detection and attribution of an anthropogenic effect on climate. *Climate Dyn.,* **12,** 77–100.

——, and Coauthors, 1996: A search for human influences on the thermal structure of the atmosphere. *Nature,* **382,** 39–46.

Tett, S. F. B., J. F. B. Mitchell, D. Parker, and M. Allen, 1996: Human influence on the atmospheric vertical temperature structure: Detection and observations. *Science,* **274,** 1170–1173.

——, T. C. Johns, and J. F. B. Mitchell, 1997: Global and regional variability in a coupled AOGCM. *Climate Dyn.,* **13,** 303–323.

Unauthenticated | Downloaded 06/13/21 11:50 PM UTC