

1 Multi-Model Assessment of Regional Surface Temperature Trends:
2 CMIP3 vs CMIP5 Historical (20C3M) Runs

3
4
5 Thomas R. Knutson, Fanrong Zeng, and Andrew T. Wittenberg

6
7 ¹Geophysical Fluid Dynamics Laboratory/NOAA, Princeton, NJ 08542

8
9 Version C; July 31, 2012

10 Submitted to *J. of Climate*

11 Email contact: Tom.Knutson@noaa.gov

12

13

14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37

Abstract.

Regional surface temperature trends from the CMIP3 and CMIP5 20th century runs are compared with observations, and assessed against a backdrop of internal climate variability as estimated from model control runs. The simulated internal climate variability is used to assess whether observed trends are “detectable” and whether the models’ historical run trends are consistent with observed trends. The trend tests focus on various periods (e.g., 1901-2010, 1951-2010, 1981-2010) and are applied at scales from global averages to individual grid points. For trends-to-2010 beginning in start years from 1901 to 1981, warming in the CMIP3 and CMIP5 simulations with volcanic forcing is consistent with observations over roughly 40-55% of the global area analyzed,. The consistent area in the CMIP5 ensemble is about 5% larger than in the CMIP3 ensemble, for trends-to-2010 that begin before 1960. The fraction of analyzed global area with no detectable trend in the observations is less than 10% for trends covering 1901-2010, but this fraction gradually grows to over 50%, and is generally slightly higher for CMIP5 than CMIP3, as the trend start date advances toward 1991. Especially for the trends beginning earlier in the record (e.g., 1901-2010) the ensemble historical run warming trend tends to be too large at lower latitudes and too small at higher latitudes. The analysis identifies regions where detection of warming trends is less robust (North Atlantic and North Pacific, the eastern tropical and subtropical Pacific), vs. areas with more robust warming signals (regions from about 40N-40S except for the eastern tropical Pacific).

38

39 **1. Introduction**

40 Are historical simulations, using climate models with the best available estimates of past climate
41 forcings, consistent with observations? This question can be examined from the viewpoint of a
42 number of different climate variables and using different comparison methods. Here we
43 compare modeled versus observed regional surface temperature trends, attempting to incorporate
44 information from a large number of climate models using various multi-model combination
45 techniques. We assess historical runs from the Coupled Model Intercomparison Project 3
46 (CMIP3; Meehl et al. 2007) and compare them with those from CMIP5 (Taylor et al. 2012).

47 The general approach used here is to compare the modeled and observed trends, in terms of both
48 magnitude and pattern, by considering trends at each gridpoint in the observational grid, as well
49 as trends over broader-scale regions. We use estimated internal climate variability, as simulated
50 in the various model control runs, to assess whether observed and simulated forced trends are
51 more extreme than those that might be expected from random sampling of internal climate
52 variability. Similarly, we use the available ensemble of simulated forced trends to assess
53 whether observed trends are compatible with the forcing-and-response hypotheses embodied by
54 those forced simulations.

55 Formal detection/attribution techniques often use a model-generated pattern from a single or set
56 of climate forcing experiments, and then regress this pattern against the observations to compute
57 a scaling amplitude (e.g., Hegerl et al. 1996; Hasselmann 1997; Allen and Tett 1999; Allen and
58 Stott 2003). If the scaling is significantly different from zero, the forced signal is detected. If
59 the scaling does not significantly differ from unity, then the amplitude of the signal agrees with

60 observations, or is at least close enough to agree within an expected range based on internal
61 climate variability. Optimal detection techniques also filter the data during the analysis such that
62 the chance of detecting a signal, if one is present in the data, is enhanced. An alternative
63 approach that is less focused on model-defined patterns has been proposed by Schneider and
64 Held (2001). In contrast to the optimal detection/attribution methods, we compare both the
65 amplitude and pattern simulated directly by the models with the observations, without rescaling
66 of patterns or application of optimization filtering. Our analysis is thus a consistency test for
67 both the amplitude and pattern of the observed versus simulated trends (e.g., Knutson et al. 1999;
68 Karoly and Wu 2005; Knutson et al. 2006). Other variants and enhancements to this general
69 type of analysis have recently been presented by Sakaguchi et al. (2012). More discussion of
70 various detection and attribution methods and their use in general is contained in Hegerl et al.
71 2009.

72 Our general approach in this study is to attempt to mimic observations with the models, in terms
73 of data coverage over time. To prevent any one model from dominating the analysis, our
74 approach attempts to weight the various models roughly equally.. Thus even if one modeling
75 center provided ten ensemble members and another only one member, or if one center provided a
76 much longer control run than the others, each of these models would still get an equal weighting.
77 (Control runs are long runs with a pre-industrial forcings that may change seasonally, but do not
78 change from year to year.) Control runs from various modeling centers are weighted equally in
79 the analysis, as long as the control run length is at least three times the length of the trend being
80 examined.

81

82 In this report, the models, methods, and observed data are described in Section 2. We examine
83 the model control runs and their variability in Section 3. Global-mean time series from the
84 20C3M historical runs are examined in Section 4. The grid point-based consistency tests are
85 presented in Section 5. Section 6 contains some additional trend analysis for data averaged over
86 larger defined regions. The discussion and conclusions are given in Section 7.

87

88 **2. Model and Observed Data Sources**

89

90 *a. Observed data*

91

92 The observed surface temperature dataset used in this study is the HadCRUT4 (Morice et al.
93 2012) which is available as a set of anomalies relative to the period 1961-1990. The dataset
94 contains some notable revisions, particularly to SSTs (HadSST3; Kennedy et al. 2011), relative
95 to previous versions, so it is important to retest earlier conclusions regarding climate trends using
96 the revised data. The dataset also contains uncertainty information, in the form of nn-ensemble
97 members sampling the estimated observational uncertainty.

98

99 To form a combined product of SST and land surface air temperature, Morice et al. (2012) adopt
100 the following procedure. If both land data and SST data are available in a particular gridbox,
101 they are weighted according to the fraction of the gridbox that is covered by land or ocean,
102 respectively. A minimum of 25% coverage is assumed, even if the fraction of the gridbox

103 covered by land is less than 25%. In our study, we use this same procedure to combine SST and
104 land surface air temperature data sets from the models we analyze. .

105

106 *b. CMIP3 and CMIP5 models*

107

108 Figure 1 displays the complete collection of models from both CMIP3 and CMIP5 used in our
109 analysis. The data were downloaded from the CMIP3 (www-pcmdi.gov/ipcc/about_ipcc.php)
110 and CMIP5 (cmip-pcmdi.llnl.gov/cmip5) model archives. We regrid the model data from
111 the 20C3M historical runs and control runs onto the observational grid. In cases where we
112 needed to use a combined the model land surface air temperature and SST data to compare with
113 observations, we used a procedure resembling that used for the observations, but using the
114 model's own land-sea mask. To mimic the data gaps in the observations, we then masked out
115 (deleted) model data at times and locations where data were labeled missing in the observations.
116 Finally, we computed the model's climatology over the same years as for observations (1961-
117 1990) and then created anomalies from this climatology. This same procedure was used for 150-
118 yr samples from the model control runs for analyses where we wanted to ensure that the control
119 runs had similar missing data characteristics to the observed data.

120

121 The forcings for the CMIP3 20C3M historical forcing runs are summarized in Rind et al. (2009;
122 Table 3.6). An important distinction among the models is the treatment of volcanic forcing. Ten
123 of the 23 CMIP3 models we examined include volcanic forcing, while 13 do not. For most of
124 our assessments, we used 19 of the CMIP3 model, of which eight included volcanic forcing. We

125 refer to these sets of models as the “Volcanic” and “Non-Volcanic” models, respectively, and
126 often distinguish between results for the two types of historical runs in our analysis. For cases
127 where we include both sets, we used the term “Volc and Non-Volc” models. All ten of the
128 CMIP5 models included in this study included volcanic forcing.

129

130 **3. Model Control Runs**

131 *a. Global mean time series*

132 The global-mean surface air temperature series from the CMIP3 and CMIP5 model control runs
133 are shown in Fig. 1. Data are displayed with arbitrary vertical offsets for visual clarity. The
134 figure also shows the observed surface temperature anomalies from HadCRUT4. The curve
135 labeled “Observed residual” was obtained by subtracting the multi-model mean of the historical
136 volcanic forcing runs. This is an estimate of the internal variability of the climate system based
137 on the residual from the estimated forcing response.

138

139 The control runs exhibit long-term drifts. The magnitude of these drifts tended to be larger in the
140 CMIP3 runs than the CMIP5 control runs, although there are exceptions. We assume that these
141 drifts are due to the models not being in equilibrium with the control run forcing, and we remove
142 these by linear trend analysis (straight lines on figure). In some CMIP3 cases the drift proceeds
143 at a given rate, but then the trend rate becomes smaller for the remainder of the run. We
144 approximate the drift in these cases with two linear trend segments, as shown in the figure, which
145 are removed to produce the drift-corrected series. The trend for these time periods is computed

146 at each model grid point and then subtracted from the model time series. One CMIP3 model
147 (IAP_fgoals1.0.g) has a strong discontinuity near year 200 of the control run. We judge this as
148 likely an artifact due to some problem with the model simulation, and we therefore chose to
149 exclude this control run from further analysis.

150

151 None of the control runs in the CMIP3 or CMIP5 samples exhibit a centennial scale trend as
152 large as the trend in the observations, aside from those with multi-century drifts as mentioned
153 above. On the other hand, the variability of observed residual series appears roughly similar in
154 scale to that from several of the control runs. Three of the CMIP3 control runs (GISS_aom,
155 GISS_model_e_h, and GISS_model_e_f) have much lower levels of variability than in the
156 observed residual series. For some sensitivity tests on the multi-model assessments, we have
157 excluded these three models to test for robustness. The Miroc_3.2_hires model also has low
158 variability, but the control runs is so short in length that it is used relatively little in our analysis,
159 since we require the control run record to be at least three times as long as the trend being
160 examined.

161 *b. Geographical distribution of variability*

162

163 The geographical distribution of the standard deviation of annual mean surface air temperature is
164 shown in Fig. 2. for CMIP3 models and Fig. 3 for CMIP5 models. These use the full available
165 time series from each control run. The time series have had the long-term drift removed as
166 discussed in section (a). The features that stand out most strongly are the enhanced variability
167 over land regions and in the eastern Equatorial Pacific. These general features (and magnitudes

168 of standard deviation) are also seen in the observations. The observed standard deviation map is
169 not shown here because of the relatively short observational record compared with the model
170 control runs, and the uncertainties in removing the forced variability component from
171 observations to create an internal variability estimate for comparison to the model control runs.
172 Versions of the control run standard deviation map which use low pass ($> X$ year) filtered data
173 (not shown) indicate that most CMIP3 and CMIP5 models have their strongest low-frequency ($>$
174 X year) variability in the polar regions and marginal sea ice areas near Antarctica, Greenland,
175 and the periphery of the Arctic Ocean.

176

177 **4. Global mean surface temperature: Historical runs**

178 *a. Time series of global mean surface temperature*

179 The global mean time series of surface temperature from the 20C3M historical runs are shown in
180 examined in Fig. 4. Thirty individual experiments using ten different models that include
181 volcanic forcing are shown in Fig. 4 (a), while 59 experiments using 23 models (with and
182 without volcanic forcing) are shown in (b). The model data series combines SST over oceans
183 and surface air temperature over land, similar to observations, and masks out periods which are
184 missing in the observed record. (All timeseries are adjusted to have zero mean in the period
185 1881-1920.)

186 The ensemble mean of the CMIP3 volcanic models (red curve in Fig. 4 (a)) agrees remarkably
187 well with observations (black curve) although the obvious volcanically induced temporary dips
188 are not in full agreement with the observed behavior for those periods. Nonetheless, one must
189 consider the role of internal climate variability in judging whether these differences are

190 significant or not. The observations are generally within the envelope of the large set of
191 individual model simulations. The spread of the individual simulations includes the model
192 uncertainty regarding the forced response, as well as internal variability generated by the models
193 (e.g., Fig. 1).

194 The combined volcanic and non-volcanic CMIP3 runs (Fig. 4 (b)) show a substantially wider
195 envelope of model behavior, as expected with the larger number of models and with the wider
196 discrepancy in forcing among the models. Since the “Non-Volcanic” runs have a substantially
197 less realistic representation of the forcing, we will generally emphasize the “Volcanic” runs in
198 panel (a) in our forced model assessments in this study.

199 *b. Spectra of global mean surface temperature*

200 Figure 5 shows the spectra of observed global mean temperature and of the individual CMIP3
201 and CMIP5 “Volcanic forcing” historical runs from Fig. 4. The enhanced power at low
202 frequencies is associated with the strong rising trend in both observations and models. At higher
203 frequencies (< 10 yr periods) the model spectra are generally within the 90% confidence
204 intervals on the observed spectral (red lines), although there is some tendency among the models
205 for lower than observed variability levels at periods less than 10 yr (frequency > 0.1 yr⁻¹).

206 Overall, the results of these comparisons suggest that the model simulations have a plausible
207 representation of variability of the climate system, in terms of the spatial pattern of variability,
208 the spectral of global mean temperature, and the direct comparison of the time series of observed
209 and historical run global mean surface temperature. These findings encourage us to use the
210 models to assess surface temperature trends at the regional scale in the following sections.

211

212 **5. Trend assessment: detection and consistency tests**

213 *a. Global means and regional “sliding trend” analysis*

214 In this section we compare the observed and simulated temperature trends to assess whether a
215 particular class of systematic temperature change (linear trend) signal has emerged from the
216 “background noise” of internal climate variability, as estimated by the models, and to assess
217 whether the observed trends are consistent with simulated trends from the historical (20C3M)
218 runs. We assess the trends across a wide “sliding range” of start years beginning in 1871. All
219 trends use 2010 as the end year. For CMIP3, we include 19 models overall in the volcanic +
220 nonvolcanic forcing results, and we include 19 models in the control run samples. Five models
221 not included in these assessments, due to drift issues, short control runs (~100 yr), or lack of a
222 necessary variable (SST) data in the archive.

223 The general procedure we use is illustrated in Fig. 6 (a) for global mean temperature. The black
224 curve in the figure shows the value of the linear trend in observed global mean temperature for
225 each beginning year from 1871-2000 and ending in the year 2010. The trend in observed
226 temperature is about 0.5°C/100 yr early in the record but has increased to over 1.5°C / 100yr by
227 around 1980. It has decreased in recent years, being near zero since 2001. The green curve
228 shows the “mean of ensemble means” for the eight CMIP3 (volcanic forcing) climate models
229 included, where each of the eight models is weighted equally, even if the modeling center
230 provided a greater than average number of within-model ensemble members.

231 The dark blue shading in Fig. 6 (a) shows the 5th to 95th percentile range of trends for the
232 corresponding window lengths from the long-term drift-adjusted control runs (Fig. 1). Each of
233 19 available CMIP3 models contributes equally to this multi-model sample, even if it has a

234 shorter control run available. We require a control run to have at least three times the data length
235 in question before it is included in our sampling, which is a random resampling technique across
236 the available data. The control data was formed into 150-yr segments with random start dates for
237 the random resampling. The 150-yr segments were then masked with the observed mask of
238 missing data over the period 1861-2010 to create data sets with similar missing data
239 characteristics to the observations. The analysis in Fig. 6 (a) shows that observed global
240 temperature trends-to-2010 of almost any length are highly unusual compared to the CMIP3
241 simulated internal variability—even for trends as short as those beginning in 1990.

242 The light pink shading in Fig. 6 (a) is a measure of the uncertainty in the CMIP3 20C3M
243 historical runs and includes the uncertainty due to different specified forcings, different forcing
244 responses, and the influence of internal variability as simulated by the models. Under an
245 assumption that internal variability in the control run is not substantially different from that in
246 the forced runs, we can use the long control run for each model to estimate the component of
247 inter-realization uncertainty that would be present in the forced trends; this is helpful, since most
248 centers did not provide enough ensemble members to precisely assess this component of the
249 uncertainty. The each randomly selected control run trend (from the eight models that also had
250 volcanic forcing runs) is combined with that model’s ensemble mean forced trend for that trend
251 length, to create a distribution of historical run trends that include the uncertainty due to internal
252 variability. The pink region is the 5th to 95th percentile range of this distribution of trends, and
253 thus relates to the uncertainty of single ensemble members (which mimics the real world, itself a
254 “single ensemble member”). In Fig. 6 (a), the black (observed) curve is always within the pink
255 shaded region, meaning that global mean temperature trends are not obviously different from the
256 CMIP3 historical run ensemble on any time scale, including for the most recent ‘weak trends’.

257 Therefore, for trends with starts through about the mid 1990s, the observed trend in global-mean
258 temperature is detectable and consistent with the CMIP3 historical runs. A similar result is
259 obtained for global mean temperature using the sample of 10 CMIP5 historical runs (Fig. 6 (b)).
260 The pink shading (uncertainty of the forced response, including internal variability) is notably
261 larger with the CMIP5 runs. Nonetheless, the observed warming is clearly detectable compared
262 to the CMIP5 control run variability distribution (blue).

263 In contrast, when the analysis is applied to the Southeast U.S. region (Fig. 6 c, d) a much
264 different result is obtained. The observed trend curve (black) rarely lies outside of the blue
265 shaded region (internal variability) meaning that except for a period from about 1950-1980 start
266 dates, the trends-to-2010 are generally not detectable in this region. In terms of consistency with
267 the model historical runs, the observed trend generally lies within the forced model ensemble
268 regions (pink shading), implying consistency for trends-to-2010 starting around 1940 and later.
269 However for start years prior to about 1940, the observations lie near the edge and often outside
270 of this 5th to 95th percentile range (pink shaded envelopes). We thus conclude that even
271 accounting for internal variability, the CMIP3 and CMIP5 historical runs trends-to-2010 tend to
272 be inconsistent or only marginally consistent with the observed surface temperature trends for
273 starting dates before about 1940. That the CMIP3 and CMIP5 models can be falsified on this
274 relatively small regional scale, means that there remain unexplained discrepancies between their
275 historical simulations and observations for trends in this region.

276 *b) Grid point-based detection and consistency assessment*

277 The above procedure can be applied to individual gridpoints and the results displayed in map
278 form. To do this, we create categories based on an observed trend's relation to the control run

279 variability (e.g., pink region in Fig. 6) and its relation to the simulated historical run trends,
280 accounting for uncertainty in the models' forced responses and internal variability. For
281 example, if the observed trend is positive and greater than the forced response (above the pink
282 region) we conclude that the trend is a "warming – detectable and greater than simulated". If the
283 observed trend is positive and lies within the pink region and outside of the blue region, we
284 conclude that the trend is "warming – detected and consistent with the simulations". If the
285 observed trend is positive, lies below the pink region and above the blue region, we conclude that
286 the trend is "warming- detectable but less than simulated. If the observed trend lies within the
287 blue region, we conclude there is "no detectable change". For cooling trends, we have analogous
288 terms to those used for the various warming cases, although these cases are relatively rare in our
289 analysis.

290

291 In Fig. 7 (a), we show the observed surface temperature linear trend map for 1901-2010. The
292 map shows warming at almost all locations. We assess this warming as highly unusual compared
293 with the CMIP3 control run (internal climate) variability over most of the global region with
294 sufficient coverage. (To determine if a grid point had "sufficient coverage" to include in our
295 maps and analyzed area, we divided a given trend period (e.g., 1901-2010) into five roughly
296 equal periods, and required that each of the five periods have at least 20% temporal coverage in
297 the monthly anomaly data.) Only in about 10% of the analyzed area (white regions in Fig. 7(c)
298 for CMIP3 and Fig. 8(c) for CMIP5) is the trend not detectable. In a very small fraction of the
299 analyzed area (less than 1% in either CMIP3 or CMIP5) is there a detectable cooling trend since
300 1901, according to our analysis.

301 Figure 7 (b) and 8 (b) show the multi-model ensemble trend maps for the CMIP3 and CMIP5
302 historical runs, weighting each of the available (volcanic) runs equally within the CMIP3 and
303 CMIP5 analyses. We used the categorization procedure described above to categorize the
304 observed vs. modeled trend comparison at each gridpoint (Figs. 7 (c); 8 (c)). The most common
305 categorization is of “warming-detected and consistent” (~40% of analyzed regions globally for
306 CMIP3 and 47% for CMIP5). The second-most common categorization is of “warming –
307 detected and greater than simulated”, which is assessed for 30% (CMIP3) and 35% (CMIP5) of
308 analyzed regions. The third-most common categorization is “warming – detected but less than
309 simulated, which is the case for about 20% (CMIP3) and 10% (CMIP5) of the area analysed.

310 In Fig. 9, we show how the percent areas that we describe above change for different start years.
311 This figure also summarizes the aggregate differences between the CMIP3 and CMIP5 results
312 (solid lines vs. dashed lines). The percent area where the warming is detected and consistent
313 with the CMIP3 or CMIP5 model stays consistently between about 40% and 55% for start dates
314 ranging from 1901 to 1981. At the same time, the percent of area with no detectable change
315 climbs steadily from 10% for 1901 start date to about 40% by 1981 start date, and reaches over
316 50% for 1991 start date. This illustrates the advantages of a long record for detectability of the
317 warming trend. The increase in percent area without a detectable trend, as one slides forward in
318 time from the 1901 start date, is compensated by a decline in the percent of area with detectable
319 warming that is either greater than or less than simulated (i.e., outside of the ‘pink envelope’ of
320 Fig. 6). The decline is largest for the classification “warming – detected and greater than
321 simulated”. Comparing the CMIP3 and CMIP5 models, the two largest differences are: CMIP5
322 has about 5% more (~40 vs. 45%) area with detectable and consistent warming than CMIP3 for
323 trends beginning in the first half of the 20th century, and about 10% less (~10 vs 20%) area with

324 “warming – detected but less than simulated” for start dates from 1901 to 1931. In short, CMIP5
325 historical runs appear at least slightly more consistent with observed trends than the CMIP3
326 historical runs are, at least for the case of trends extending from the early 20th century to 2010.
327 There is slightly less area with detectable warming trends according to the CMIP5 models,
328 particularly for trends-to-2010 beginning from 1931 start date on.

329 The corresponding maps for 1951-2010 and 1981-2010 observed trends, ensemble mean
330 historical run trends, and the categorization maps for those trends for the CMIP3 and CMIP5
331 models are shown in Figs. 10- 13 (panels a-c). These show the general spatial patterns
332 associated with the changes in trend behavior for different start dates and for the CMIP3 and
333 CMIP5 historical runs noted above. The loss of detectability, as one proceeds to mid-20th
334 century start dates, occurs first in the extratropical North Atlantic (north of 40°N) and over large
335 parts of the North Pacific, extending into the tropics, as seen for the 1951-2010 trends (Figs. 10
336 c, 11 c). For the late 20th century start dates (e.g., 1981-2010; Fig. 12c, 13c) the region of no
337 detectable warming expands to cover most of the southern oceans, south of 40°S, and extending
338 south from 20°S in the South Atlantic. This region also expands to include most of the eastern
339 tropical and subtropical Pacific and much of the northern extratropics over Eurasia, North
340 America, and the North Pacific. Tropical and subtropical regions within about 40-50 degrees of
341 the equator (except for the eastern Pacific) are generally the regions with still a detectable (and
342 generally consistent) warming signal, for trends beginning as late as 1981.

343
344 The remaining panels (d-n) in Figs. 7, 8, 10-13 show classification maps for the observed vs.
345 historical runs, but in this case the metric is percentage of individual CMIP3 or CMIP5 models
346 that are classified with the particular category for that geographic location and beginning year of

347 the trend (all ending in 2010). That is, the determination of whether a given CMIP3 or CMIP5
348 individual model is included in a category (e.g., “warming- detectable and consistent”) is based
349 on the evaluation of the historical runs and control runs for that model alone. The most
350 consistent signals across the models are for the “warming – detectable” category, which has all
351 or almost all models in that category across large areas of the globe for 1901-2010 trends, and
352 even for much of the tropics and subtropics for the relatively recent trends (1981-2010). A
353 notable distinction between the CMIP3 and CMIP5 results in these figures is in the maps of
354 fraction of models that are consistent with observed trends, including non-detectable changes.
355 The CMIP5 model ensemble has considerably higher average fraction than the CMIP3 ensemble.
356 Figure 14 shows a summary statistic for the individual models. In this figure we compare the
357 fraction of analyzed area where there is both a detectable change and where the change is
358 consistent with the individual climate model. Note that this metric does not include the fraction
359 of area where a climate model is consistent with observations but there is not a detectable trend.
360 While all metrics have shortcomings, this particular metric has at least some compensation
361 effects, where the enhancement of consistency due to increased internal variability is partly
362 compensated by a reduction in the area with detectable trends for models with increased internal
363 variability. We plan to explore other metric approaches that explore this parameter space more
364 thoroughly in future extensions of this work.

365 The results in Fig. 14 show that the individual CMIP3 and CMIP5 models have rather similar
366 behavior in terms of fraction of area with consistent detectable trends. There is somewhat more
367 spread among the CMIP5 models (although there are more models in the sample as well.) This
368 metric tends to reach a peak value around 1960-1970 start date before declining for later start
369 dates.

370 **6. Extensions and Applications of the Analysis**

371 The analysis presented in this study introduces a framework for trend analysis that has many
372 possible applications and extensions. Several of these, which are either planned, in progress, or
373 completed. However, we cannot include these here as there are too many figures which do not
374 fit within the length constraints of the journal. These extensions are briefly introduced here. We
375 are creating a web site based largely on this analysis which will contain a growing collection of
376 figures that will provide access to many of these extensions and applications as they become
377 available. These are briefly discussed below.

378 a. Sensitivity analyses

379 A number of questions could be posed about our analysis, such as what do the plots look like for
380 individual seasons, what if we had used 97th and 2.5th percentiles instead of 95th and 5th, what if
381 we had left certain “low variability” models (Section 3a) out of the analysis, what if we had used
382 a different observed data set or observed ensembles from the HadCRUT4 data product and so
383 forth. Some of these sensitivity analyses have already been completed and are available on the
384 above web site.

385 b. Focus on individual regions

386 Figure 15 shows a number of regions for which we have prepared extensive trend analyses like
387 that in Fig. 6. We have done these analyses for various 4-month seasons, using CMIP3 or
388 CMIP5 models, using 97.5th and 2.5th percentiles, leaving out certain CMIP3 control runs with
389 lower variability levels, and other sensitivity tests. The plots are too numerous to present in this
390 paper, but are accessible on the above web site.

391 c. Focus on individual models

392 Figures similar those in this multi-model analysis can also be prepared for individual models in
393 the CMIP3 and CMIP5 archive. We are in the process of producing these. These analyses may
394 be of interest as feedback to the individual centers and to others interested in individual model
395 characteristics. The results, as they are updated, will be posted to the website above.

396

397 d. Weighting of future projections

398 Figure 14 shows an example of evaluation of individual models in terms of the fraction of global
399 analyzed area with trends-to-2010 that consistent with observations. This analysis suggests a
400 means of weighting future projections from different models based on the models' levels of
401 agreement with past trends as in Fig. 14. As mentioned above, there are ways of trying to create
402 improved model comparison metrics, which we plan to explore in a future study.

403 e. Application to Other Variables

404

405 An extension of this methodology would be explore application to other climate variables such
406 as precipitation. We are planning to do this, beginning with precipitation, in upcoming work and
407 to report on these developments in a future study as well as through updates and extensions of
408 these on the above web site.

409

410

411 **7. Summary and Conclusions**

412 The purpose of this analysis has been to introduce and apply a framework for assessing regional
413 surface temperature trends from the CMIP3 and CMIP5 models using a multi-model sampling
414 approach. We showed the behavior of the various control runs of the CMIP3 and CMIP5
415 models. We used the control run variability to help assess whether observed trends were unusual
416 or not compared with control run (internally generated) variability. We also used the control run
417 variability to help assess whether observed trends were consistent with (or alternatively,
418 significantly different from) trends from the historical (20C3M) simulations. In the separate
419 CMIP3 and CMIP5 analyses, we generally attempt to give different models equal weight, even
420 when a modeling center provides fewer ensemble members or shorter control runs. Tests are
421 applied at global and regional scales, as well as at individual grid points on the observed data
422 grid where there is sufficient data coverage over the period of the trend. Results are summarized
423 using classification maps and global percent area statistics.

424 Our analysis of variability (standard deviation maps, spectral analysis, and time series
425 inspection) suggest that the CMIP3 and CMIP5 models provide a plausible representation of
426 internal climate variability, with some likely exceptions which were noted for some models and
427 regions.

428 The assessment of the trends allowed us to identify regions where the detection of warming
429 trends is most robust (in terms of still being detectable, according to the models, for relatively
430 late start dates, such as 1981). These regions tend to be in the tropics and subtropics, but outside
431 of the eastern Pacific, which is influenced by strong interannual variability associated with
432 ENSO. The reduced global area with detectable trends as one examines later start dates for trends
433 in the record (all trends ending in 2010) illustrates the advantages of long records for trend
434 detection in the context of this model-based assessment. The analysis also suggests a modestly

435 closer agreement of models with observed trends for CMIP5 models compared to CMIP3
436 models—at least for the relatively longer trends-to-2010 that begin in the first half of the 20th
437 century.

438 For trends-to-2010 beginning from the early 20th century, about 40-50% of the analyzed regions
439 globally have a detectable warming that is consistent with the 20C3M historical runs, with
440 slightly higher percentage for the CMIP5 simulations. The fraction of area with no detectable
441 change is only about 10% for trends 1901-2010, but increases steadily to over 50% as the
442 beginning year is moved forward to 1981. The fraction of area with detectable and consistent
443 warming stays relatively constant for start years through about 1981, before falling below 40%
444 for trends from 1991-2010. The “loss” of detectable warming regions as one moves forward
445 with the start dates, is mainly a “loss” in regions with detectable warming that is inconsistent
446 with the historical runs, which decreases from about 50% for 1901-2010 to less than 10% for
447 trends 1991-2010. That is, for the most recent trends (1991-2010), the trends are classified
448 predominantly as either non-detectable relative to the control runs, or as detectable warming that
449 is consistent with model historical runs (for both CMIP3 and CMIP5 models). The shorter the
450 epoch, the larger the contribution of internal variability to the trend, leading to a greater spread
451 (uncertainty) for sampled trends.

452 As has been noted in a previous paper using a similar methodology with two climate models
453 (Knutson et al. 2006), disagreement between modeled and observed trends in this type of
454 analysis can occur due to shortcomings of models (internal variability simulation; response to
455 forcing), shortcomings of the specified specified historical forcings, or problems with the
456 observed data. The HadCRUT4 data set (Morice et al. 2012) contains multiple ensemble
457 members that attempt to characterize the uncertainties in the observations. We have performed

458 some preliminary tests using these ensembles to assess the spread of observed trend estimates.
459 These tests thus far indicate that even at the regional scale, the spread in trend estimates due to
460 observational uncertainties, as contained in the ensembles, is generally much smaller than the
461 spread in model simulated trends due to both internal variability and differences in forced
462 responses in the historical runs (e.g., Fig. 6). However, it is possible that other observational
463 datasets could have somewhat different trends.

464 We have attempted to at least partially address the issue of uncertainties in the simulation of
465 internal climate variability and in the response to historical forcing by using multi-model
466 ensembles. Nonetheless, the CMIP3 and CMIP5 simulations represent an “ensemble of
467 opportunity” which cannot necessarily be expected to represent the true structural uncertainty in
468 results, due to shortcomings/uncertainties in the models and climate forcings. From a different
469 perspective, Shin and Sardeshmukh (2011) have noted that the CMIP3 models do not simulate
470 historical trends of temperature and precipitation as realistically as atmospheric models forced by
471 observed trends in tropical SSTs—a problem they attribute to model errors as opposed to climate
472 noise (internal variability). Clearly there appears to scope for improvement in model simulations
473 of past trends using historical forcings.

474 While these issues lack a final resolution, the methodology shown here can at least help to
475 quantify the uncertainties associated with the climate change detection problem. The results
476 show that when CMIP3 and CMIP5 historical runs are confronted with observed surface
477 temperature variations and trends, across a wide range of trend start dates and at various
478 geographical locations around the globe, warming is found that is generally much more
479 consistent with forced simulations than with unforced simulations. This provides further support
480 for the claim of a discernable influence of humans on climate, via anthropogenic forcing agents

481 like increased greenhouse gases. A future enhancement of these findings would be to compare
482 the CMIP5 all-forcing historical runs with runs that include only natural forcings, to provide a
483 more direct assessment of the roles of anthropogenic versus natural forcings in observed
484 temperature trends at the regional scale.

485

486 Acknowledgments. We thank the Met Office Hadley Centre and the Climatic Research Unit,
487 Univ. of East Anglia, for making the HadCRUT4 data set available to the research community.
488 We thank the modeling groups participating in CMIP3 and CMIP5, and PCMDI for generously
489 making the model output used in our report available to the community.

490

491

492

493

494 **References**

495 Allen, M. R., and P. A. Stott, 2003: Estimating signal amplitudes in optimal fingerprinting. Part
496 I: Theory. *Clim. Dyn.*, 21, 477-491.

497 Allen, M. R., and S.F.B. Tett, 1999: Checking for model consistency in optimal fingerprinting.
498 *Clim. Dyn.*, 15, 419-434.

499 Hasselmann, K., 1997: Multi-pattern fingerprint method for detection and attribution of climate
500 change. *Clim. Dyn.*, 13, 601-612.

- 501 Hegerl, G. C., et al. 2009: Good practice guidance paper on detection and attribution related to
502 anthropogenic climate change. Available from IPCC: [www.ipcc.ch/pdf/supporting-](http://www.ipcc.ch/pdf/supporting-material/ipcc_good_practice_guidance_paper_anthropogenic.pdf)
503 [material/ipcc_good_practice_guidance_paper_anthropogenic.pdf](http://www.ipcc.ch/pdf/supporting-material/ipcc_good_practice_guidance_paper_anthropogenic.pdf)
- 504 Hegerl, G.C., H. v. Storch, K. Hasselmann, B. D. Santer, U. Cubasch, and P. D. Jones, 1996:
505 Detecting greenhouse gas induced climate change with an optimal fingerprint method. *J.*
506 *Climate*, 9, 2281-2306.
- 507 Karoly, D.J., and Q. Wu, 2005: Detection of regional surface temperature trends. *J. Clim.*, **18**,
508 4337–4343.
- 509 Kennedy, J. J., N. A. Rayner, R. O. Smith, D. E. Parker, and M. Saunby, 2011: Reassessing
510 biases and other uncertainties in sea-surface temperature observations measured in situ
511 since 1850, part 2: biases and homogenization. *J. Geophys. Res.*, 116, D14104,
512 doi:10.1029/2010JD015220.
- 513 Knutson, T.R., T.L. Delworth, K.W. Dixon, and R.J. Stouffer, 1999: Model assessment of
514 regional surface temperature trends (1949-1997). *J. Geophys. Res.*, **104**, 30981–30996.
515
- 516 Knutson, T.R., et al., 2006: Assessment of twentieth-century regional surface temperature trends
517 using the GFDL CM2 coupled models. *J. Clim.*, **19**, 1624–1651.
518
- 519 Meehl, G. A. et al., 2007: The WCRP CMIP3 multimodel dataset: A new era in climate change
520 research. *Bull. Amer. Meteor. Soc.* **88**, 1383–1394.
521

- 522 Morice, C. P., J. J. Kennedy, N. A. Rayner, and P. D. Jones, 2012: Quantifying uncertainties in
523 global and regional temperature change using an ensemble of observation estimates: The
524 HadCRUT5 data set. *J. Geophys. Res.*, 117, D08101, doi:10.1029/2011JD017187.
- 525 Rind, D., M. Chin, G. Feingold, D. Streets, R. A. Kahn, S. E. Schwartz, and H. Yu, 2009:
526 Modeling the effects of aerosols on climate. In *Atmospheric Aerosol Properties and*
527 *Climate Impacts, A Report by the U.S. Climate Change Science Program and the*
528 *Subcommittee on Global Change Research*. [Mian Chin, Ralph A. Kahn, and Stephen E.
529 Schwartz (eds.)]. National Aeronautics and Space Administration, Washington, D.C.,
530 USA.
- 531 Sakaguchi, K. X. Zeng, and M. A. Brunke, 2012: Temporal- and Spatial-scale dependence of
532 three CMIP3 climate models in simulating the surface temperature trend in the twentieth
533 century. *J. Climate*, 25, 2456-2470.
- 534
- 535 Santer, B. D., T. M. L. Wigley, and P. D. Jones, 1993: Correlation method in fingerprint
536 detection studies. *Clim. Dyn.*, 8, 265-276.
- 537
- 538 Schneider, T., and I.M. Held, 2001: Discriminants of twentieth-century changes in Earth surface
539 temperatures. *J. Clim.*, **14**, 249–254.
- 540
- 541 Shin, S.-I., and P. D. Sardeshmukh, 2011: Critical influence of the pattern of tropical ocean
542 warming on remote climate trends. *Clim. Dyn.*, 36, 1577-1591.
- 543

544 Taylor, K.E., R.J. Stouffer, and G.A. Meehl, 2012: An overview of CMIP5 and the experiment
545 design. *Bull. Amer. Meteor. Soc.*, **93**, 485-498.

546

547

548

549

550 **Figure Captions**

551

552 Fig. 1. Timeseries of global mean annual mean surface air temperature (2 m) anomalies from the
553 CMIP3 (a, b) and CMIP5 (c) preindustrial control runs. Observed global mean surface
554 temperature (HadCRUT4, combining SST and land surface air temperature anomalies) is also
555 shown on the diagrams for comparison. The curves labeled “Observed residual” or “HadCRU4
556 residual” were created by subtracting the multi-model ensemble mean surface temperature (from
557 masked SSTs and land surface air temperatures from the 20C3M historical runs for either CMIP3
558 or CMIP5) from the observed temperature. Straight lines (one or two segments) through the
559 control run time series depict the long term linear drift. The long term drift over these years is
560 calculated at each grid point and then subtracted from the model control run series before
561 performing further analysis in our study. The various curves have been displaced vertically by
562 arbitrary constants for visual clarity.

563

564 Fig. 2. Standard deviation ($^{\circ}\text{C}$) of annual mean surface air temperature from the CMIP3 pre-
565 industrial control runs (e.g., Fig. 1 a,b). The long term linear drifts (periods identified by the
566 linear line segments in Fig. 1 a,b) were removed prior to computing the standard deviation. The
567 individual plots are labeled with the name of the model/center and classified as “Non-V” (non-
568 volcanic) or “V” (volcanic) depending on whether than model’s historical run used in this study
569 included volcanic forcing or not. Note that the control runs on which the figure are based do not
570 have episodic volcanic forcing and have been masked for observed missing data periods. The
571 final panel (“obs”) is an observational estimate of internal variability of SST (oceanic regions)

572 and surface air temperature (land regions) constructed by removing the CMIP3 eight-model
573 ensemble (Volcanic models) estimate of the forced climate response from the observed
574 temperature record over 1949-2010.

575

576 Fig. 3. As in Fig. 2 but for the 10 CMIP5 models analyzed in this study. The final panel (“obs”)
577 is an observational estimate of internal variability of SST (oceanic regions) and surface air
578 temperature (land regions) constructed by removing the CMIP5 ten-model ensemble (Volcanic
579 models) estimate of the forced climate response from the observed temperature record over
580 1949-2010.

581 Fig. 4. Timeseries of global mean surface temperature anomalies (combined SST and land
582 surface air temperature) from observations (HadCRUT4; black curves) and CMIP3 (a, b) or
583 CMIP5 (c) 20C3M historical runs (orange curves) in degrees Celsius. The historical runs in (b)
584 include 23 CMIP3 models with and without volcanic forcing (as in Fig. 1 (a,b) but excluding
585 IAP_FGOALS1.0_g). Those in (a) are from CMIP3 models with volcanic forcing. All of the
586 CMIP5 model runs shown in (c) included volcanic forcing. The red curves show the multi-
587 model ensemble means, which was computed by weighting each model equally (as opposed to
588 each individual model run equally). All series have been re-centered so that the mean value for
589 the years 1881-1920 is zero. Model data were masked with the observed temporally evolving
590 missing data mask.

591

592 Fig. 5. Variance spectra as a function of frequency for observed global mean surface
593 temperature (combined SST and land surface air temperature), in black with 90% confidence

594 intervals shown in red, plotted against spectra for the individual (a) CMIP3 and (b) CMIP5
595 “Volcanic forcing” historical runs (green) based on the time series in Fig. 4. The spectra in (c)
596 and (d) are based on observed or model historical runs where the multi-model ensemble surface
597 temperature from the 20C3M volcanically forced historical runs is subtracted from the observed
598 global mean temperature series to form a residual. Similarly, this multi-model ensemble (either
599 CMIP3 or CMIP5) is subtracted from each individual historical run to form a modeled residual
600 for comparison to the observed.

601

602 Fig. 6. Trends (deg C/100 yr) in surface temperature as a function of starting year, with all
603 trends ending in 2010, for the CMIP3 (a,c) and CMIP5 (b,d) models. The black curves are from
604 observations (HadCRUT4). The green curves are the multi-model ensemble means, with each
605 model weighted equally. The blue shading shows the 5th to 95th percentile range of trends of the
606 given length based on random resampling of the model control runs, with each model sampled
607 equally frequently regardless of control run length. The pink shading shows the range obtained
608 by using the same control run samples as for the blue shading, but adding onto each control run
609 trend the ensemble mean trend, from the given start year, of that model’s all forcing run. Violet
610 shading shows where the pink and blue shaded regions overlap. Region used: Global (a,c) or
611 the Southeast United States (b,d), with boundaries of the latter region shown in Fig. 15. The 19
612 CMIP3 models used here and in subsequent assessment figures include all listed in Fig. 1 except
613 IAP_FGOALS1.0_g, INGV_ECHAM4, MIROC3.2_hires, MIP_ECHAM5, and NCAR PCM1.

614

615 Fig. 7. Geographical distribution of: (a) HadCRUT4 observed or (b) CMIP3 multi-model
616 (volcanic models) ensemble mean surface temperature trends (1901-2010) in degrees C per 100
617 yr. The observed trend is assessed in terms of the multi-model ensemble mean trends and
618 variability in (c). In (c) the meaning of the different colors is shown to the right of the color
619 scale. Panels (d-h) show the fraction of the 10 individual CMIP3 models whose historical
620 forcing (including volcanic) runs meet the criteria listed below the panel. The criteria are: d)
621 detectable cooling that is more than simulated; e) detectable cooling that is consistent with the
622 model; f) detectable cooling that is less than simulated; g) no detectable change; h) detectable
623 warming that is less than simulated; i) detectable warming that is consistent with the model; j)
624 detectable warming that is more than simulated; k) detectable warming (sum of h,i,j); l)
625 detectable warming that is consistent or greater than simulated (i+j); m) observed and simulated
626 trends are consistent (including non-detectable changes that are consistent); and n) observed and
627 simulated trends are inconsistent (1-m).

628 Fig. 8. As in Fig. 7, but for the ten CMIP5 models analyzed in the study.

629 Fig. 9. Summary assessment of trends-to-2010 comparing the CMIP3 (solid lines) and CMIP5
630 (dashed lines) multi-model ensembles (historical 20C3M runs with volcanic forcing). The
631 fraction of global analyzed areas meeting certain criteria (see graph labels) are shown as a
632 function of start year.

633 Fig. 10. As in Fig. 7, but for trends over the period 1951-2010.

634 Fig. 11. As in Fig. 7, but trends over the period 1951-2010 for the ten CMIP5 models analyzed
635 in the study.

636 Fig. 12. As in Fig. 7, but for trends over the period 1981-2010.

637 Fig. 13. As in Fig. 7, but trends over the period 1981-2010 for the ten CMIP5 models analyzed
638 in the study.

639 Fig. 14. Individual a) CMIP3 and b) CMIP5 models are assessed for consistency with observed
640 surface temperature trends-to-2010 for start years from 1901 to 1991. Plotted is the percent of
641 analyzed global area where each model's (legend) multi-member ensemble mean forced trends
642 are consistent (accounting for internal variability) with the observed trends. The trends are
643 analyzed at each grid point where there is sufficient temporal data coverage for the trend in
644 question (see text).

645 Fig. 15. Map illustrating regions where trend analyses (like those in Fig. 6, but with additional
646 augmented analyses as discussed in the text) are available online (web site).

647

648

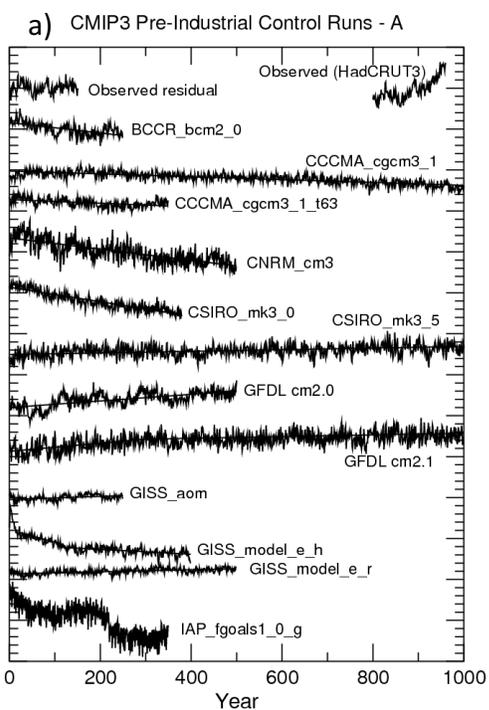
Fig. 1

649

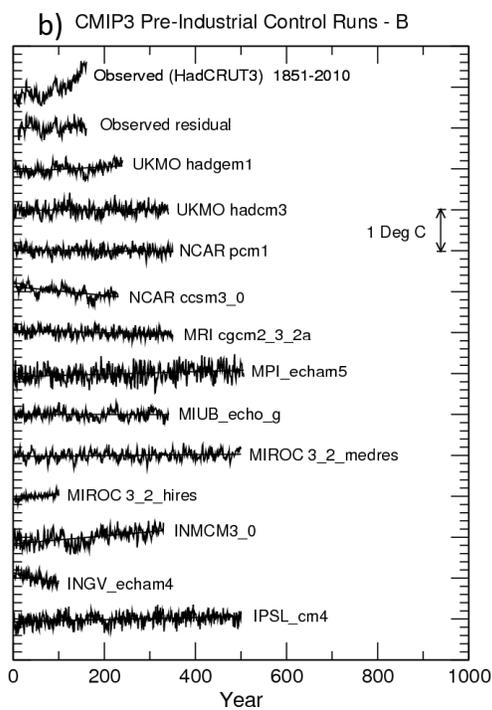
650

651

Global Surface Air Temperature Anomalies



Global Surface Air Temperature Anomalies



c) CMIP5 Pre-industrial Control runs

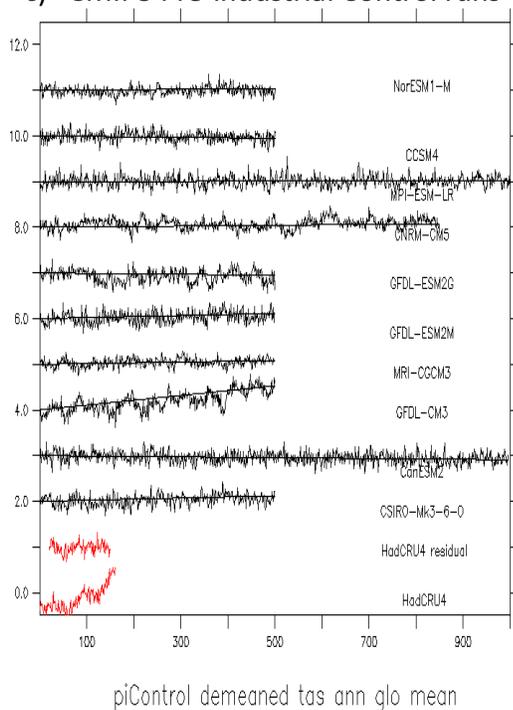


Fig. 1. Timeseries of global mean annual mean surface air temperature (2 m) anomalies from the CMIP3 (a, b) and CMIP5 (c) preindustrial control runs. Observed global mean surface temperature (HadCRUT4, combining SST and land surface air temperature anomalies) is also shown on the diagrams for comparison. The curves labeled “Observed residual” or “HadCRU4 residual” were created by subtracting the multi-model ensemble mean surface temperature (from masked SSTs and land surface air temperatures from the 20C3M historical runs for either CMIP3 or CMIP5) from the observed temperature. Straight lines (one or two segments) through the control run time series depict the long term linear drift. The long term drift over these years is calculated at each grid point and then subtracted from the model control run series before performing further analysis in our study. The various curves have been displaced vertically by arbitrary constants for visual clarity.

Fig. 2

652

653

654

655

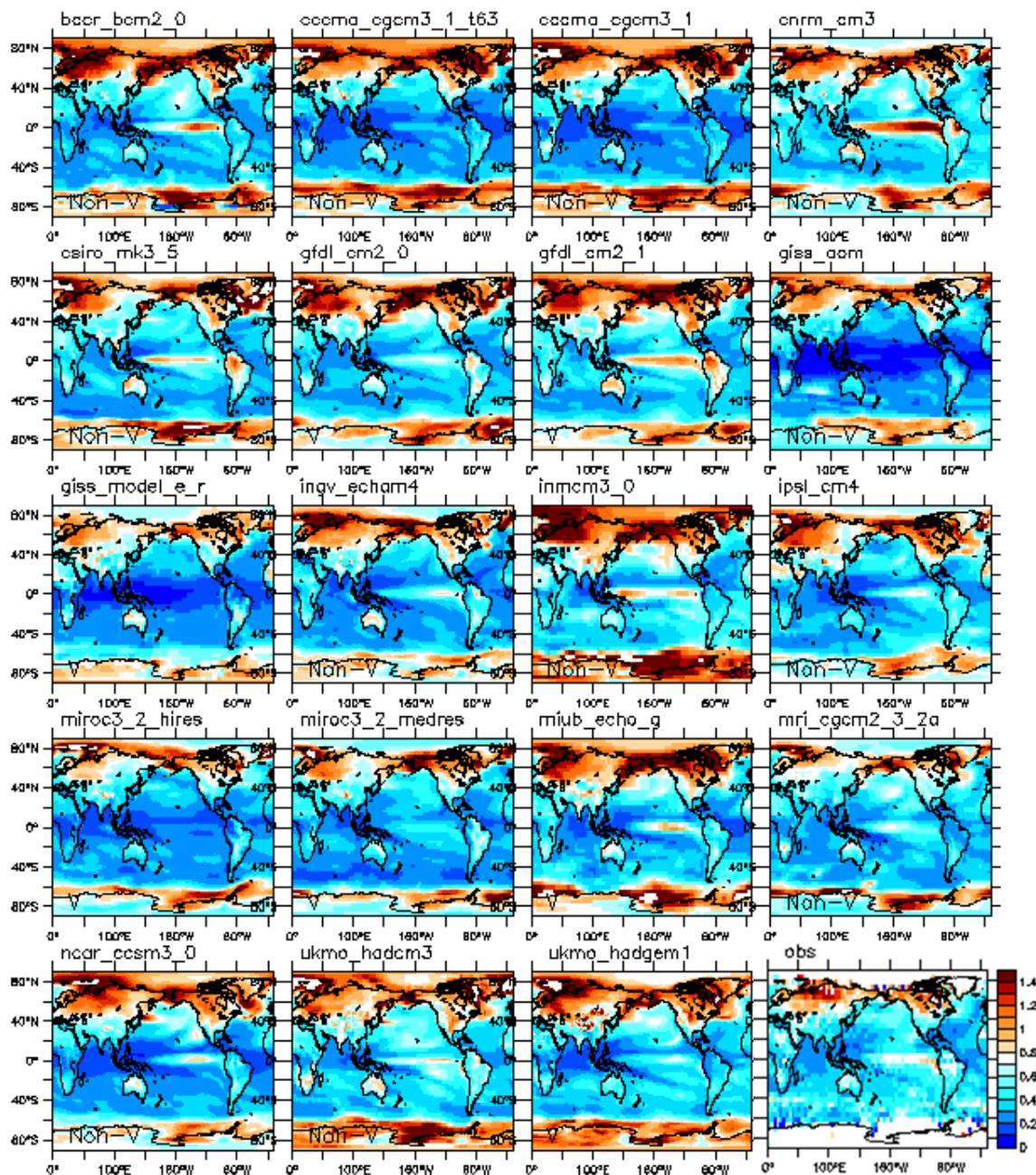


Fig. 2. Standard deviation ($^{\circ}\text{C}$) of annual mean surface air temperature from the CMIP3 pre-industrial control runs (e.g., Fig. 1 a,b). The long term linear drifts (periods identified by the linear line segments in Fig. 1 a,b) were removed prior to computing the standard deviation. The individual plots are labeled with the name of the model/center and classified as “Non-V” (non-volcanic) or “V” (volcanic) depending on whether than model’s historical run used in this study included volcanic forcing or not. Note that the control runs on which the figure are based do not have episodic volcanic forcing and have been masked for observed missing data periods. The final panel (“obs”) is an observational estimate of internal variability of SST (oceanic regions) and surface air temperature (land regions) constructed by removing the CMIP3 eight-model ensemble (Volcanic models) estimate of the forced climate response from the observed temperature record over 1949-2010.

Fig. 3

656

657

658

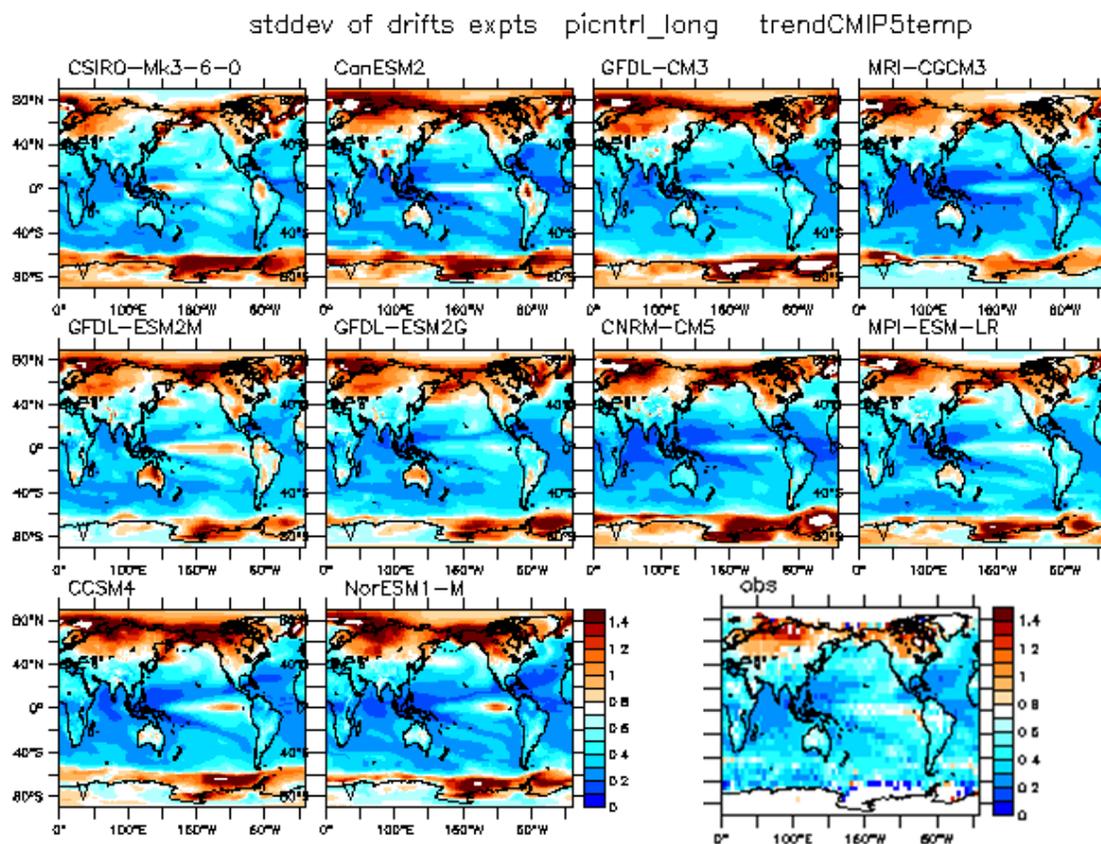


Fig. 3. As in Fig. 2 but for the 10 CMIP5 models analyzed in this study. The final panel (“obs”) is an observational estimate of internal variability of SST (oceanic regions) and surface air temperature (land regions) constructed by removing the CMIP5 ten-model ensemble (Volcanic models) estimate of the forced climate response from the observed temperature record over 1949-2010.

659

660

661

662

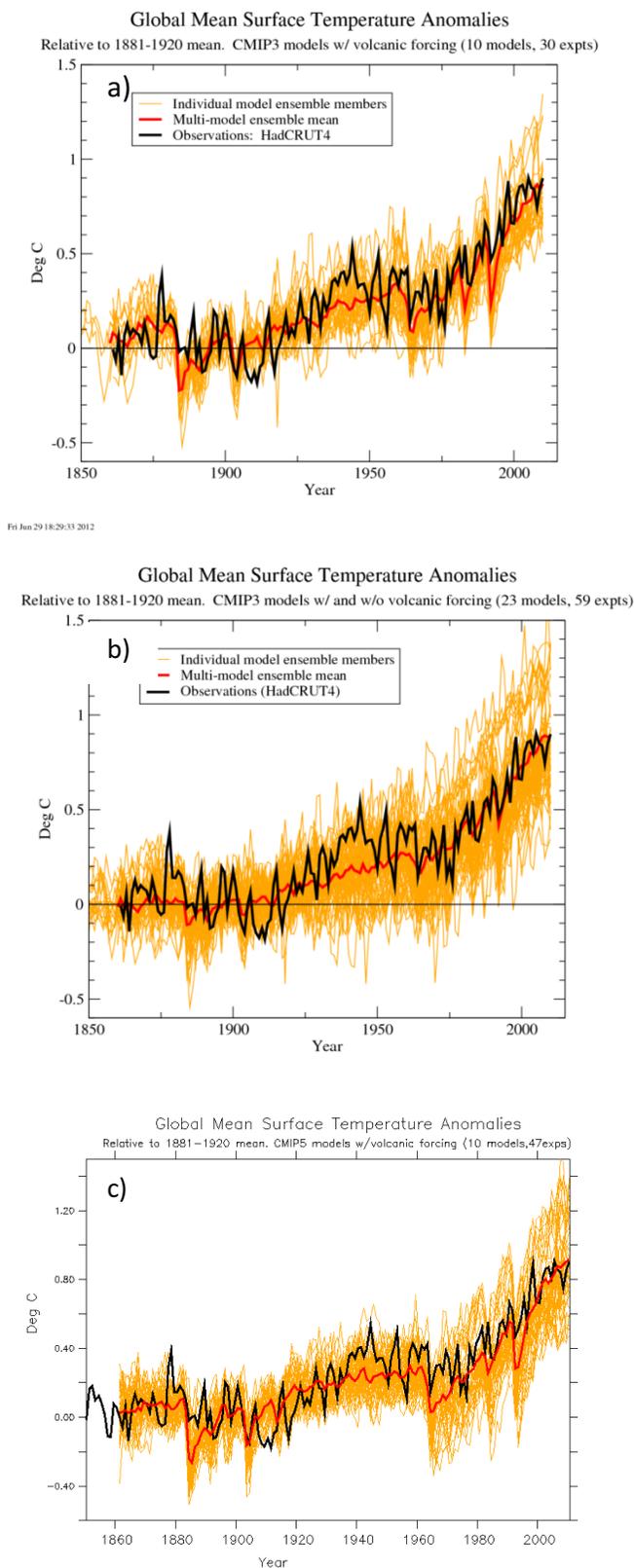


Fig. 4

Fig. 4. Timeseries of global mean surface temperature anomalies (combined SST and land surface air temperature) from observations (HadCRUT4; black curves) and CMIP3 (a, b) or CMIP5 (c) 20C3M historical runs (orange curves) in degrees Celsius. The historical runs in (b) include 23 CMIP3 models with and without volcanic forcing (as in Fig. 1 (a,b) but excluding IAP_FGOALS1.0_g). Those in (a) are from CMIP3 models with volcanic forcing. All of the CMIP5 model runs shown in (c) included volcanic forcing. The red curves show the multi-model ensemble means, which was computed by weighting each model equally (as opposed to each individual model run equally). All series have been re-centered so that the mean value for the years 1881-1920 is zero. Model data were masked with the observed temporally evolving missing data mask.

663

Fig. 5

664

665

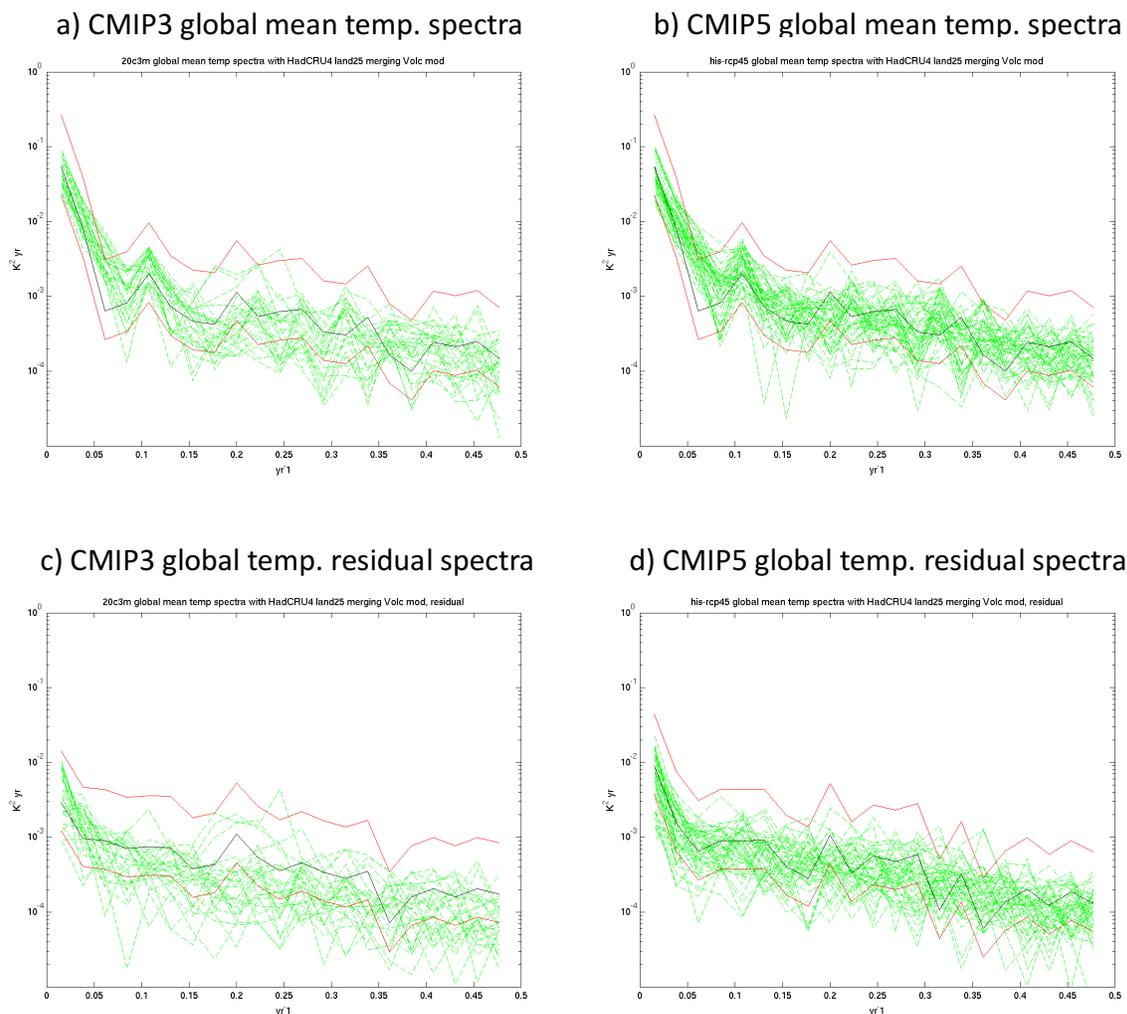
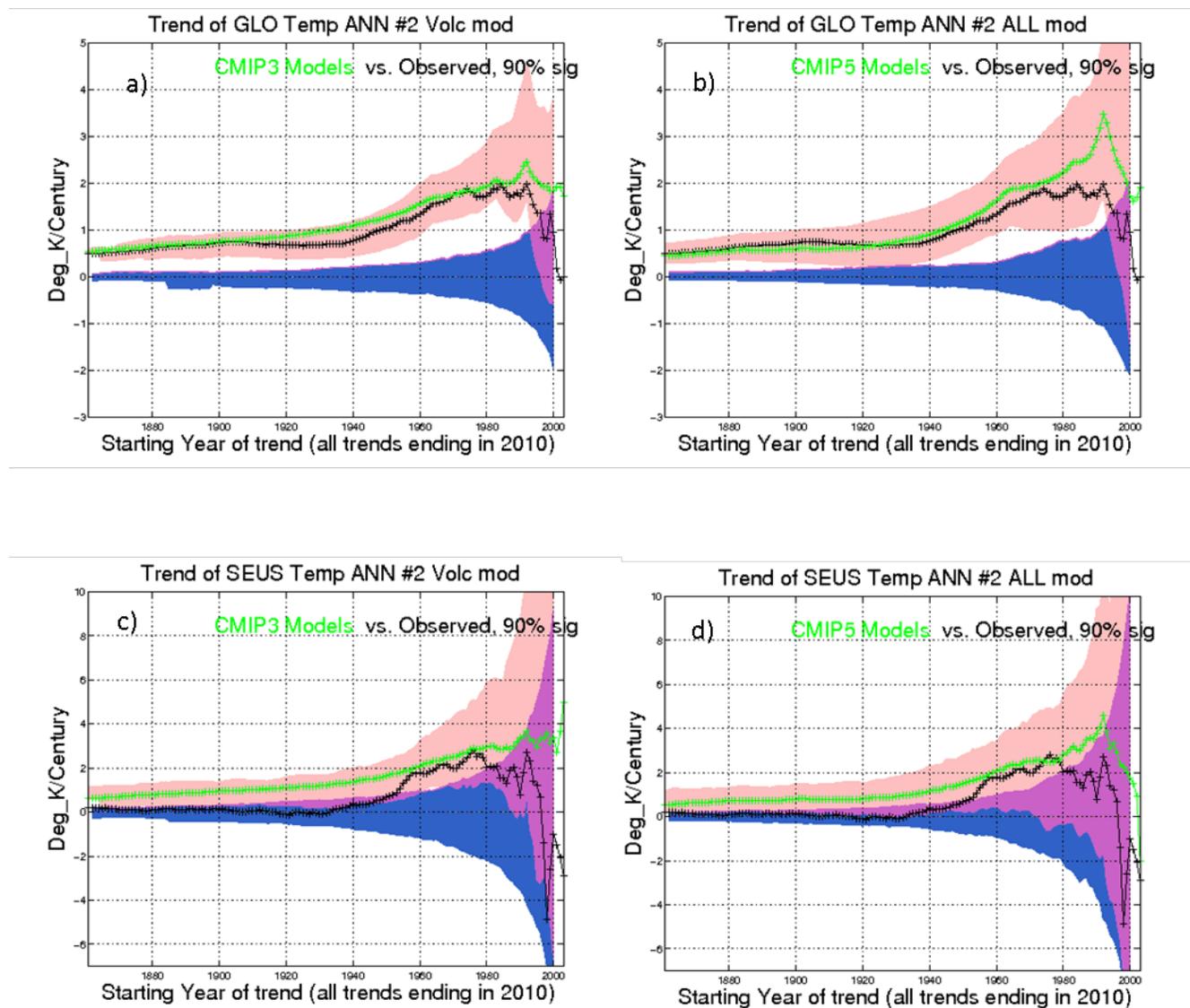


Fig. 5. Variance spectra as a function of frequency for observed global mean surface temperature (combined SST and land surface air temperature), in black with 90% confidence intervals shown in red, plotted against spectra for the individual (a) CMIP3 and (b) CMIP5 “Volcanic forcing” historical runs (green) based on the time series in Fig. 4. The spectra in (c) and (d) are based on observed or model historical runs where the multi-model ensemble surface temperature from the 20C3M volcanically forced historical runs is subtracted from the observed global mean temperature series to form a residual. Similarly, this multi-model ensemble (either CMIP3 or CMIP5) is subtracted from each individual historical run to form a modeled residual for comparison to the observed.

Fig. 6



666
 667 Fig. 6. Trends (deg C/100 yr) in surface temperature as a function of starting year, with all trends
 668 ending in 2010, for the CMIP3 (a,c) and CMIP5 (b,d) models. The black curves are from observations
 (HadCRUT4). The green curves are the multi-model ensemble means, with each model weighted
 equally. The blue shading shows the 5th to 95th percentile range of trends of the given length based on
 random resampling of the model control runs, with each model sampled equally frequently regardless of
 control run length. The pink shading shows the range obtained by using the same control run samples
 as for the blue shading, but adding onto each control run trend the ensemble mean trend, from the given
 start year, of that model's all forcing run. Violet shading shows where the pink and blue shaded regions
 overlap. Region used: Global (a,c) or the Southeast United States (b,d), with boundaries of the latter
 region shown in Fig. 15. The 19 CMIP3 models used here and in subsequent assessment figures
 include all listed in Fig. 1 except IAP_FGOALS1.0_g, INGV_ECHAM4, MIROC3.2_hires,
 MIP_ECHAM5, and NCAR PCM1.

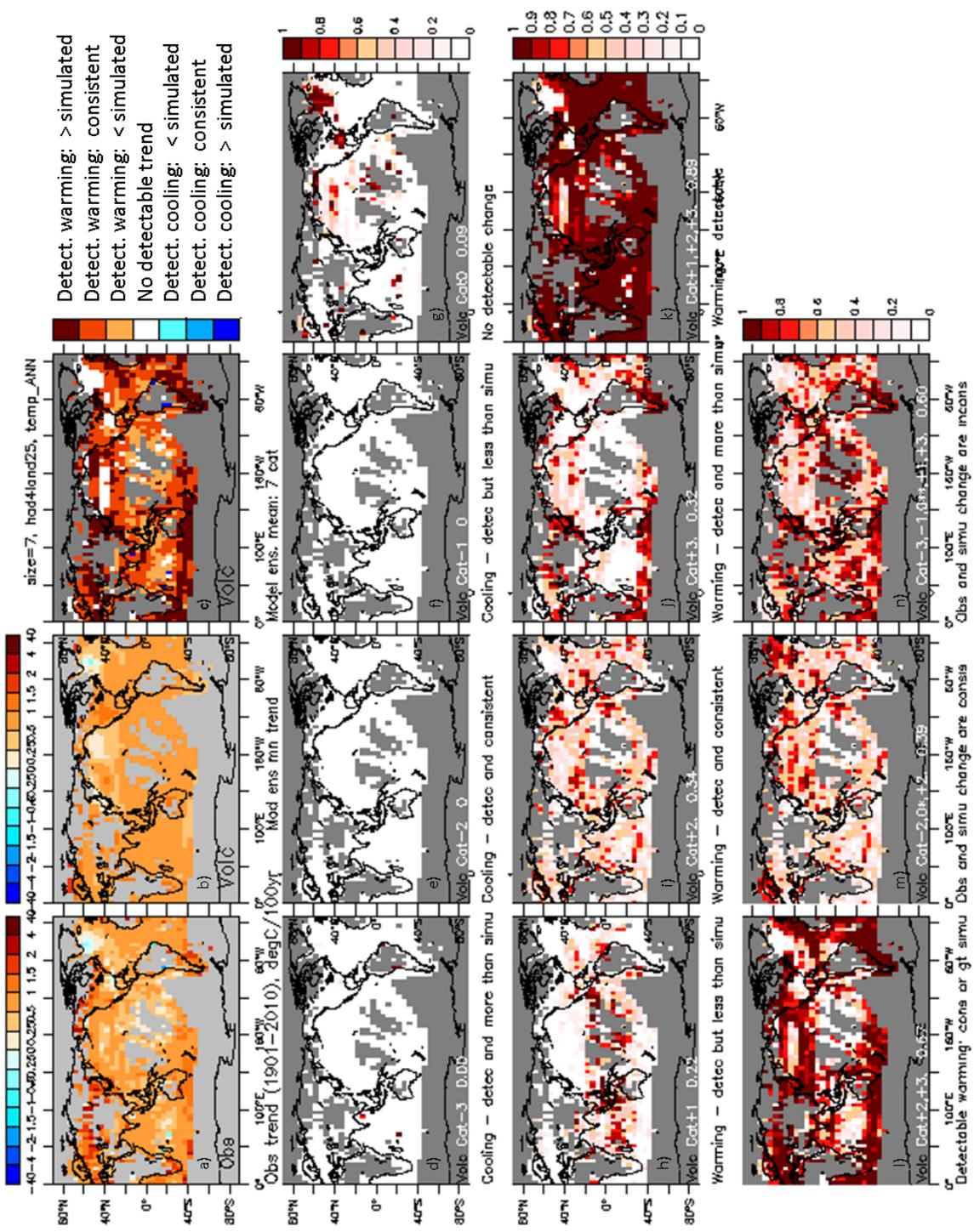


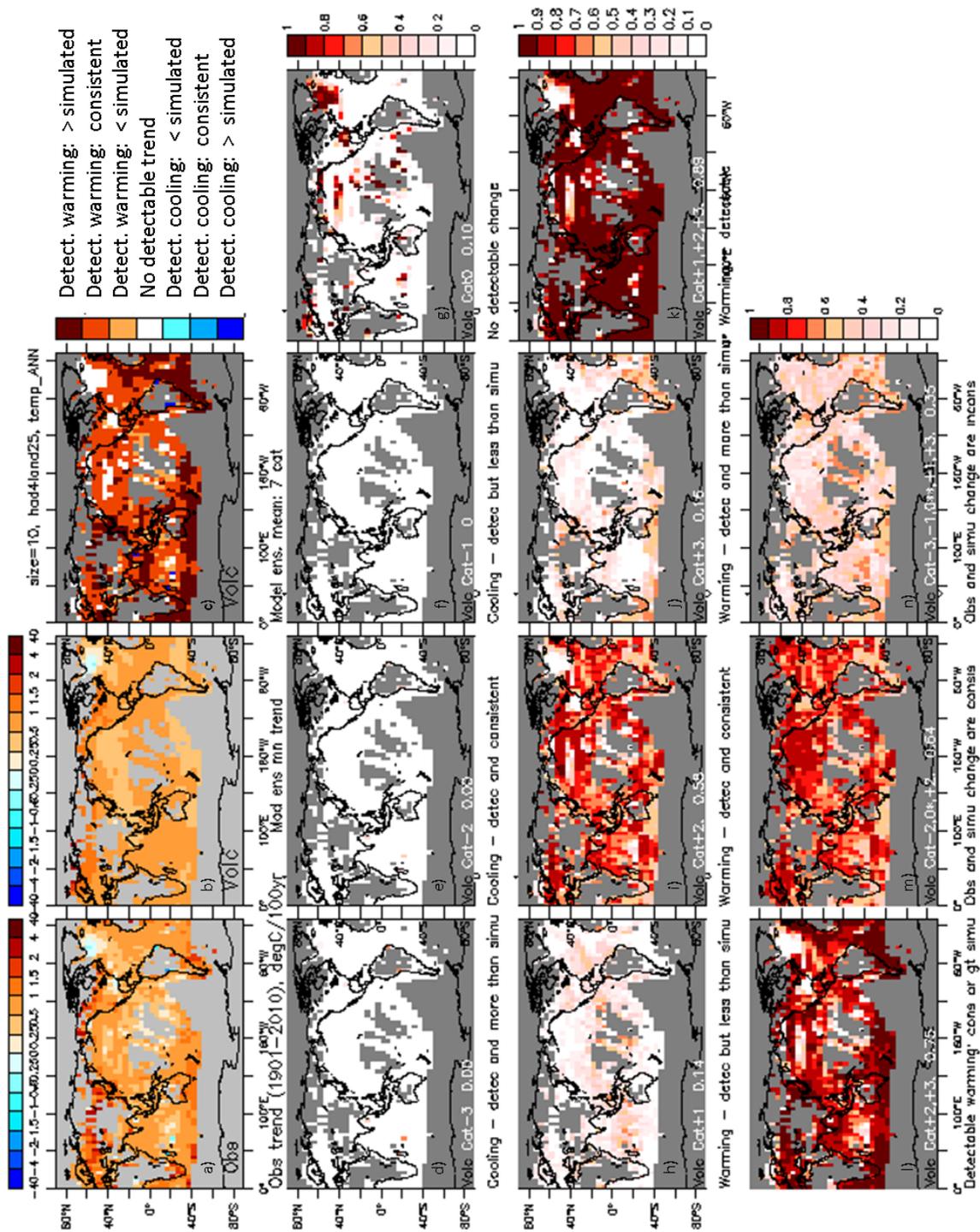
Fig. 7

673

674

Fig. 7. Geographical distribution of: (a) HadCRUT4 observed or (b) CMIP3 multi-model (volcanic models) ensemble mean surface temperature trends (1901-2010) in degrees C per 100 yr. The observed trend is assessed in terms of the multi-model ensemble mean trends and variability in (c). In (c) the meaning of the different colors is shown to the right of the color scale. Panels (d-h) show the fraction of the 10 individual CMIP3 models whose historical forcing (including volcanic) runs meet the criteria listed below the panel. The criteria are: d) detectable cooling that is more than simulated; e) detectable cooling that is consistent with the model; f) detectable cooling that is less than simulated; g) no detectable change; h) detectable warming that is less than simulated; i) detectable warming that is consistent with the model; j) detectable warming that is more than simulated; k) detectable warming (sum of h,i,j); l) detectable warming that is consistent or greater than simulated (i+j); m) observed and simulated trends are consistent (including non-detectable changes that are consistent); and n) observed and simulated trends are inconsistent (1-m).

Fig. 8



675
676
677

Fig. 8. As in Fig. 7, but for the ten CMIP5 models analyzed in the study.

Fig. 9.

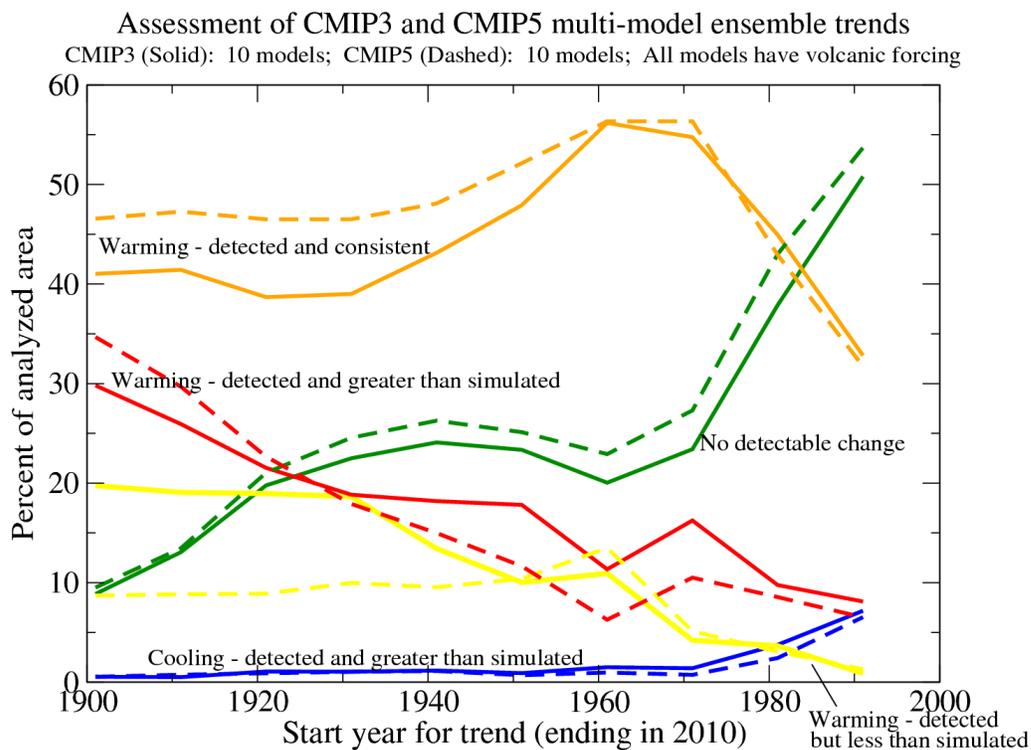
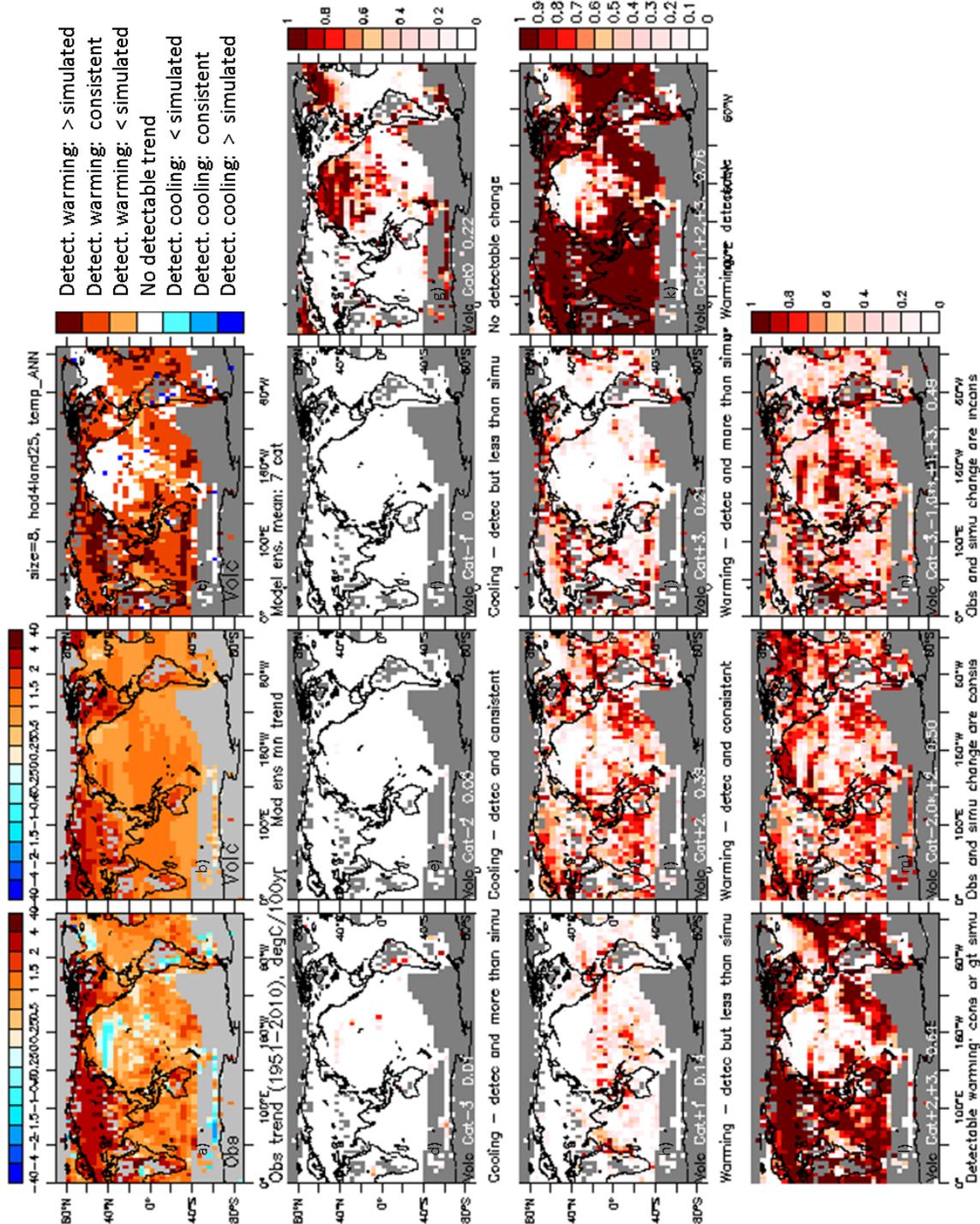


Fig. 9. Summary assessment of trends-to-2010 comparing the CMIP3 (solid lines) and CMIP5 (dashed lines) multi-model ensembles (historical 20C3M runs with volcanic forcing). The fraction of global analyzed areas meeting certain criteria (see graph labels) are shown as a function of start year.

Fig. 10

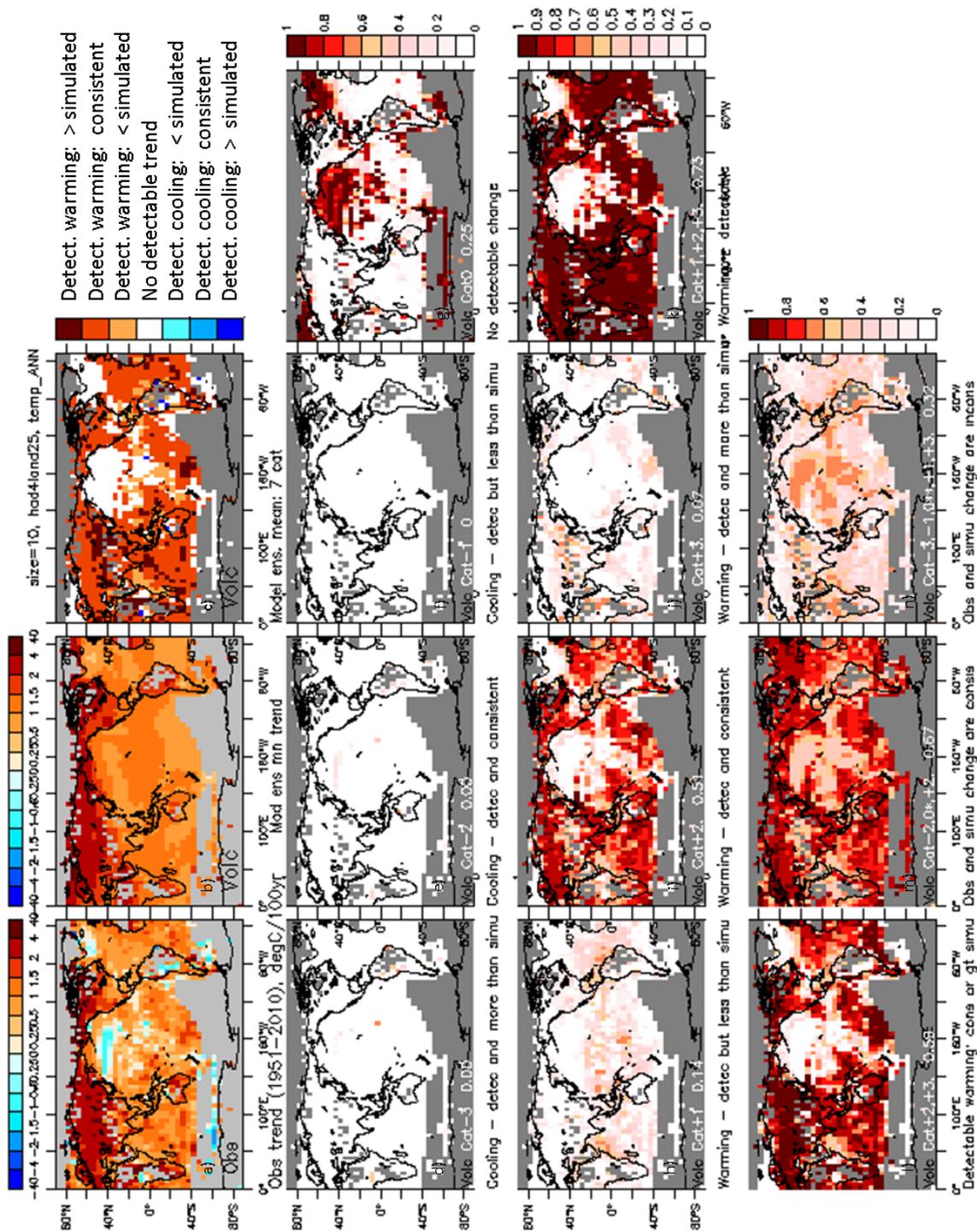


679

Fig. 10. As in Fig. 7, but for trends over the period 1951-2010.

680

Fig. 11



681

Fig. 11. As in Fig. 7, but trends over the period 1951-2010 for the ten CMIP5 models analyzed in the study.

Fig. 12

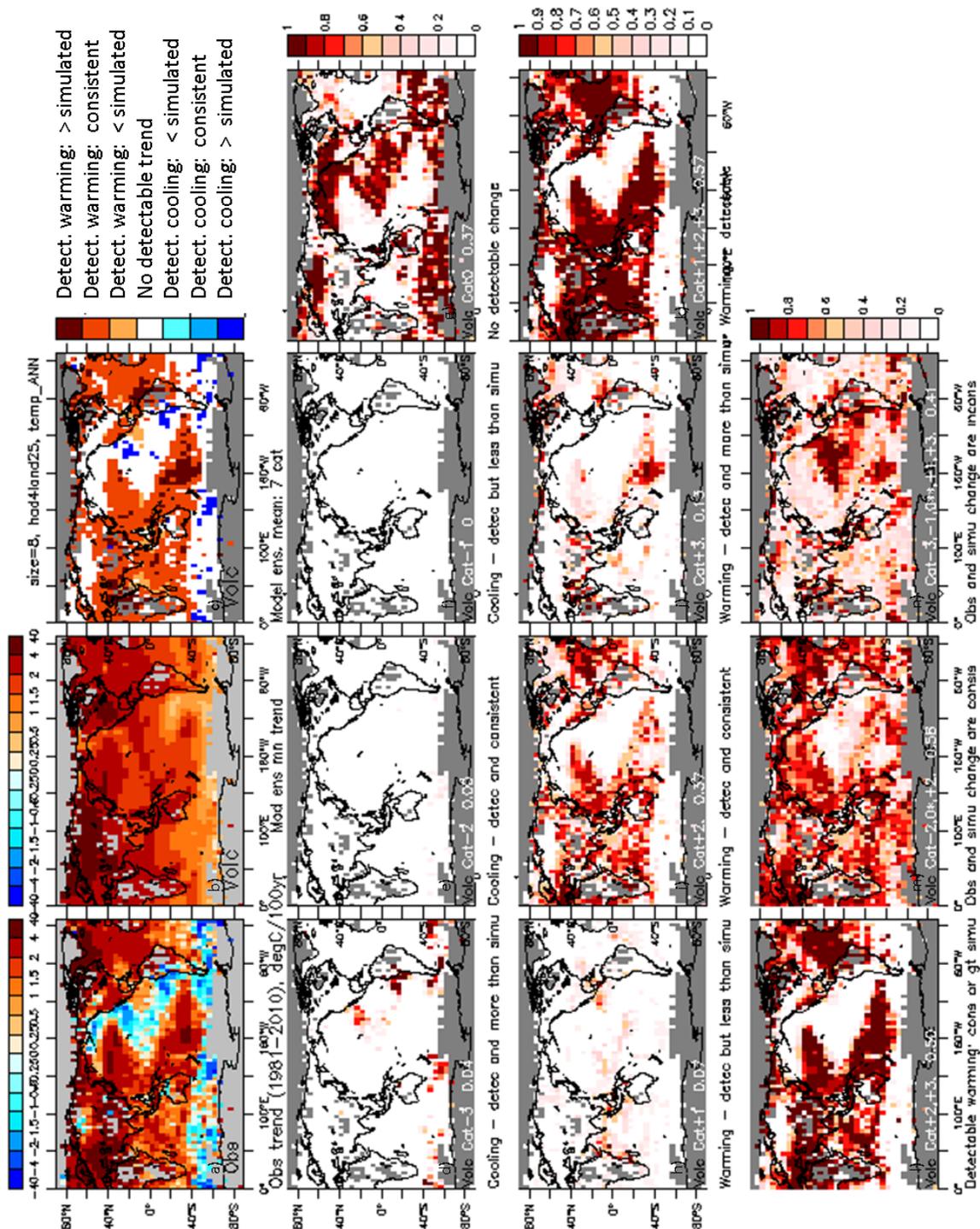


Fig. 12. As in Fig. 7, but for trends over the period 1981-2010.

Fig. 13

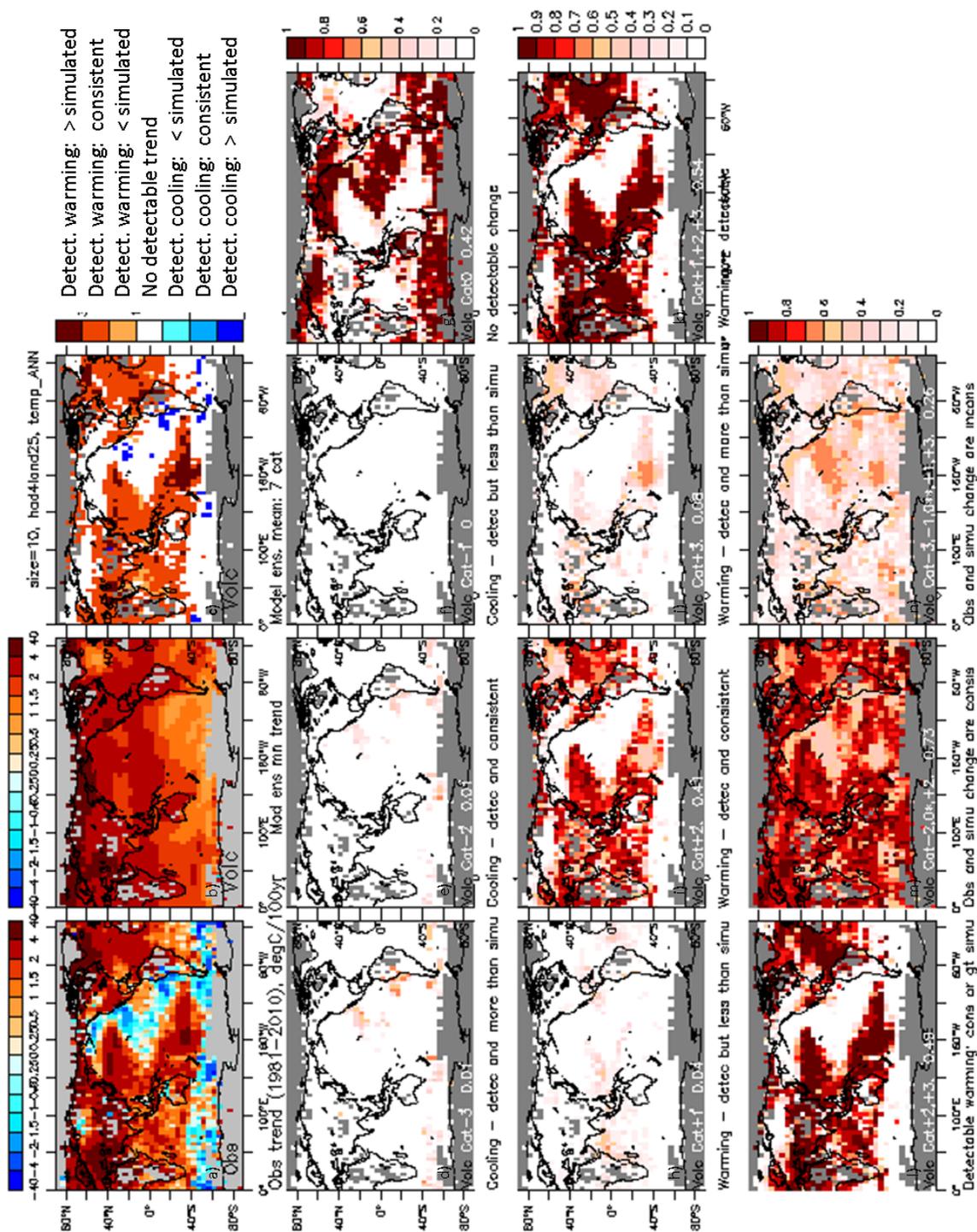


Fig. 13. As in Fig. 7, but trends over the period 1981-2010 for the ten CMIP5 models analyzed in the study.

684

685

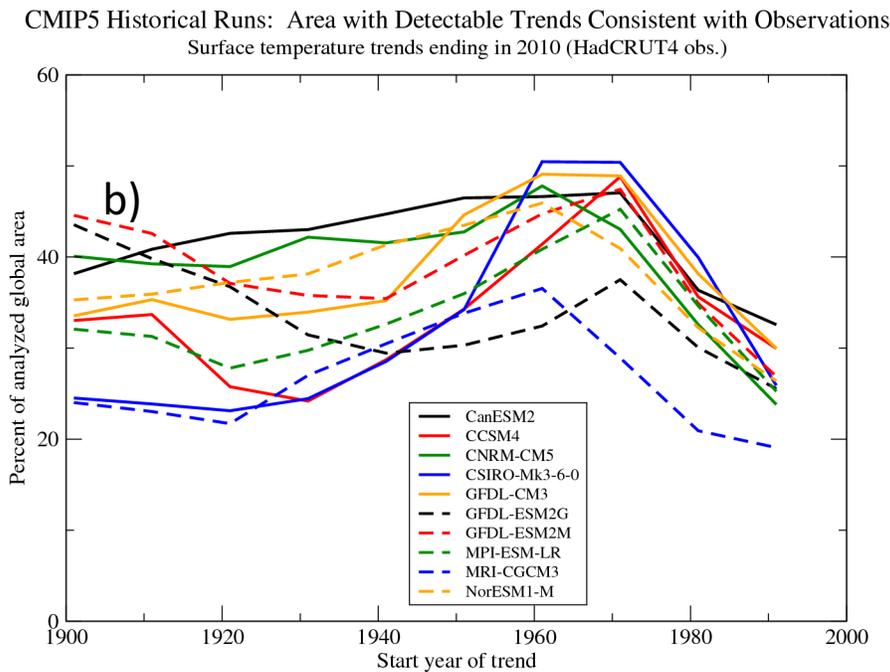
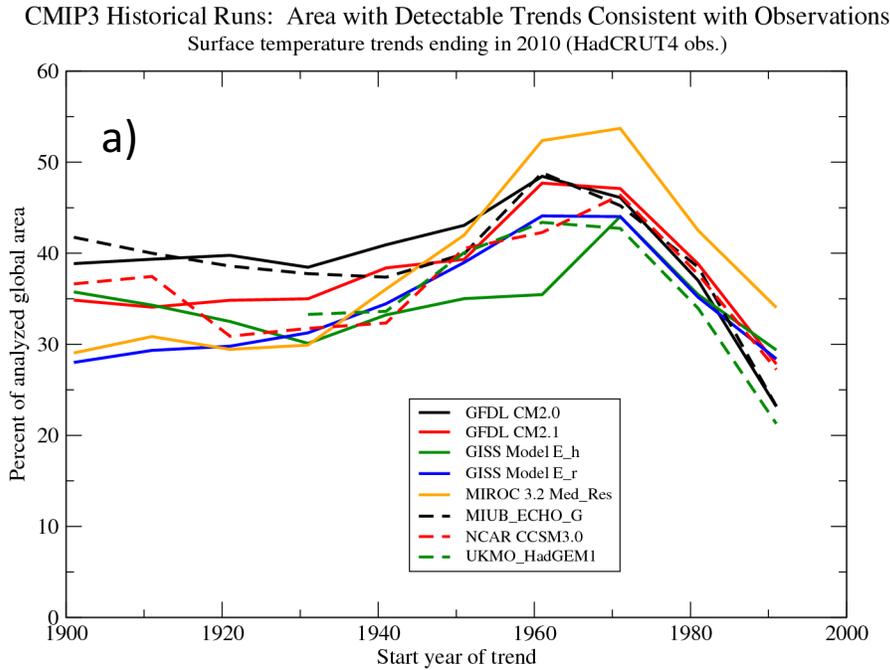


Fig. 14. Individual a) CMIP3 and b) CMIP5 models are assessed for consistency with observed surface temperature trends-to-2010 for start years from 1901 to 1991. Plotted is the percent of analyzed global area where each model's (legend) multi-member ensemble mean forced trends are consistent (accounting for internal variability) with the observed trends. The trends are analyzed at each grid point where there is sufficient temporal data coverage for the trend in question (see text).

686

687

Fig. 15

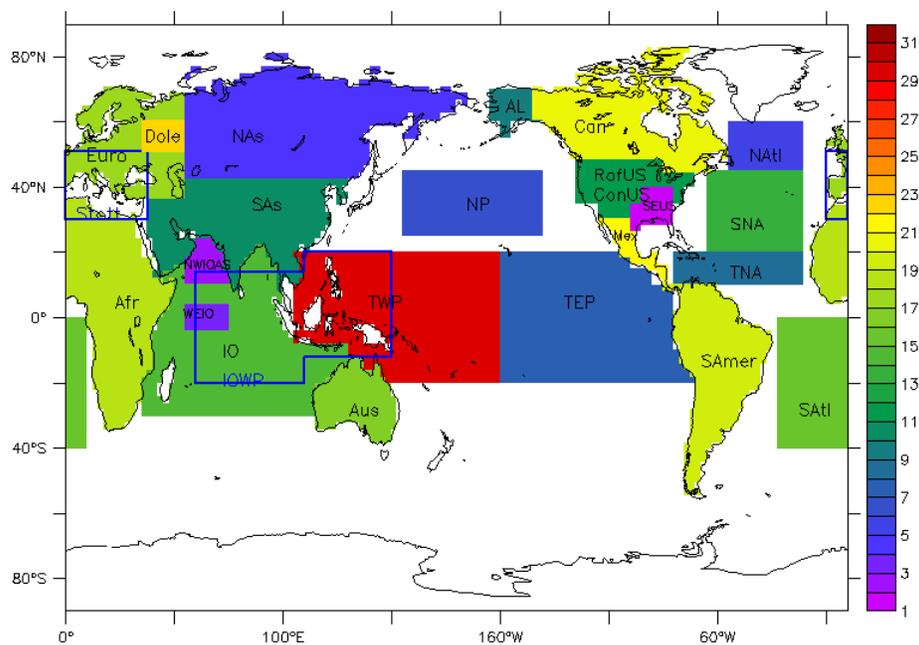


Fig. 15. Map illustrating regions where trend analyses (like those in Fig. 6, but with additional augmented analyses as discussed in the text) are available online (web site).