Investigating the recent apparent hiatus in surface temperature increases: 1. Construction of two 30-member Earth System Model ensembles

Stephen Outten1, Peter Thorne1,2, Ingo Bethke3, and Øyvind Seland4

1Nansen Environmental and Remote Sensing Centre, Bjerknes Centre for Climate Research, Bergen, Norway, 2Department of Geography, Maynooth University, Maynooth, Ireland, 3UniResearch Climate, Bjerknes Centre for Climate Research, Bergen, Norway, 4Norwegian Meteorological Institute, Oslo, Norway

Abstract: The recent Intergovernmental Panel on Climate Change report, along with numerous studies since, has suggested that the apparent global warming hiatus results from some combination of natural variability and changes to external forcings. Herein the external forcings for greenhouse gases (GHG), long-lived trace gases, volcanic and tropospheric aerosols, and solar irradiance have been replaced in the Norwegian Earth System Model using recent observational estimates. The potential impacts of these alternative forcings, and by extension the internally generated variability, are examined through two 30-member ensembles covering the period 1980 to 2012. The Reference ensemble uses the Coupled Model Intercomparison Project phase 5 (CMIP5) historical forcings extended with the Representative Concentration Pathway 8.5 (RCP8.5) scenario, while the Sensitivity ensemble uses the alternative forcings. Over the hiatus period defined here as 1998–2012, all of the forcings show a change between the Sensitivity and Reference experiments and have a combined net forcing change of 0.03 W m⁻². The GHG forcing is 0.012 W m⁻² higher in the Sensitivity forcings. The alternative solar forcing differs from the Reference forcing by 0.008 W m⁻², the same as the alternative volcanic forcing that was based on the latest estimates from NASA Goddard Institute for Space Studies. Anthropogenic aerosol emissions were replaced using the EU-Eclipse4.1 data set and produce a mean forcing change of 0.11 W m⁻² over the period. Part I details the creation of the two 30-member ensembles and the characterization for various time series and particular relevance to the explanation of the hiatus. A detailed investigation of the two resulting ensembles global surface temperature behavior is given in Part 2, along with comparisons to observational data sets.

1. Introduction

The global mean surface temperature has increased more slowly over the past 15 years than over the last 50 years. The trend has fallen from 0.11 K decade⁻¹ over 1951–2012 to 0.04 K decade⁻¹ over 1998–2012 (Flato et al., 2013), although updates to observational estimates since Cowtan and Way, 2014; Karl et al., 2015) change these numbers somewhat for some observational estimates. Even so, in all current observational data sets, the period 1998–2012 exhibits a slow rate of global mean surface temperature warming than the preceding 15 years. While this apparent pause of hiatus in the global mean warming represents the mean global annual surface temperature trends, it has been shown to have both a seasonal and geographical distribution, with the largest decrease in trends found over the continents of the Northern Hemisphere during the winter (Cohen et al., 2012). The global warming hiatus has drawn great attention, in part because it has not been reproduced by the vast majority of global climate models when forced by future climate projections based on IPCC AR5. The 4th assessment report (AR4) of the Intergovernmental Panel on Climate Change (IPCC) stated that the hiatus is likely due to a combination of low internal variability in the climate system and a reduced trend in the external forcings (i.e., emissions and efficient climate response). While these findings have been mostly consistent with recent studies (e.g., Huber and Knutti, 2014; Kosaka and Xie, 2013; Meehl et al., 2014; Schmidt et al., 2014; Manabe and Stainforth, 2015), these new existences of hypotheses in the literature regarding the different factors contributing to the global warming hiatus. It is important to stress that several climate models show a different response in the absence of internal variability and the potential causes of an internal variability such as Pacific Ocean heat content (Meehl et al., 2014) or the El Niño–Southern Oscillation (ENSO) (Kebbi et al., 2014) which may be important.

One of the prevailing hypotheses regarding internal variability in the climate system was put forward by Kosaka and Xie [2013]. They perform a coupled model experiment in which they prescribed the sea surface

© 2015. The Authors. This is an open access article under the terms of the Creative Commons Attribution 3.0 Unported License. This is an open access article under the terms of the Creative Commons Attribution 3.0 Unported License, which permits use, distribution, and reproduction in any medium, provided the original work is properly cited. In this case, the work is properly cited.
temperature (SST) in the Eastern Pacific from historical anomalies. The results highlighted the role of Pacific variability and ENSO, such as the strong El Niño event in 1998 and the La Niña-like decadal cooling, in explaining much of the cooling that is the cause of the hiatus. Therewere built on the findings of Kosaka and Xie [2013] by examining climate models based on their representation of ENSO [Risbey et al., 2014], by examining changes in winds that drive the cooling for Pacific variability [England et al., 2014], and by adjusting winds in a global climate model (GCM) to reproduce the cooling effect [Watanabe et al., 2014]. One caveat to the findings of Kosaka and Xie [2013] is that the model produced strong warming over the winter in Russia in contradiction to the strong cooling that has been observed [Cohen et al., 2012] and that is a part of the hiatus.

The observed Russian winter cooling has been linked to the reduction of sea ice in the Russian sector of the Arctic, which causes strong local warming and the formation of a thermocline. This results in a decrease of the meridional temperature gradient between the Arctic and midlatitudes and a weakening of the westerly jet that transports heat and moisture from the Atlantic to the continent [Dutton and Bau, 2012]. At the same time, the large-scale flow is changed, facilitating the transportation of cold air from the Siberian Arctic down to the midlatitudes of Eurasia [Petoukhov and Semenov, 2010; Mor et al., 2014]. It is also unclear to what extent this Russian cooling is a predictable secondary response to anthropogenic forcing or is attributable to natural variability. Nor is it obvious whether such cooling would continue into the future under transient climate change.

While internal variability is likely partially responsible for detemining the global warming hiatus, the other potential part of the explanation lies in changes to the external forcings, including greenhouse gases (GHG— including ozone-depleting substances), volcanic and anthropogenic aerosols, solar forcing, ozone, and the effect of stratospheric water vapor. The inclusion of these forcings into the global climate models (GCMs) has, for the past 20 years, been coordinated through the Coupled Model Intercomparison Project (CMIP), the 5th phase of which, CMIP5, has formed the basis of the models assessed within IPCC AR5. In CMIP5, the observed forcings are applied only until the end of the historical experiments period in 2005. Beyond this, the forcings are based on the Representative Concentration Pathway (RCP) scenarios. Further on, due to model run times and delays in assimilating the latest observations, the few years in mediate preceding 2005 may also have questionable forcings. Hence, the CMIP5 models only include accurate forcing estimates for the last 5 to 7 years of the hiatus period, defined here as 1998–2012. Numerous individual studies have investigated how a slowdown in global warming could be the result of changes to a single set of forcings, e.g., non-CO2 GHGs [Hansen et al., 2000], ozone-depleting substances [Estrada et al., 2013], volcanic [Rilley et al., 2014; Solomon et al., 2011; Santer et al., 2014], and anthropogenic aerosols [Neely et al., 2013].

At least two preceding studies have considered the sensitivity to forcing mispecification in climate models using specified scenarios specifically [Schmidt et al., 2014; Santer et al., 2014]. Santer et al. [2014] showed that better specification of volcanic eruptions produced a 15% decrease in the differences between observations and simulations of surface temperature. What is missing to date is an aggregated approach in which the impact of the individual forcings is included simultaneously in fully coupled runs. An investigation was made in this vein by utilizing approximations of how the forcings have changed, i.e., linear interpolations for GHGs and ozone-depleting substances, estimated aerosol impacts, estimated effects of volcanic aerosols, etc., and adjusting the CMIP5 ensemble with a correction based on a simple impulse/response model [Schmidt et al., 2014].

As part of an assessment on how changes in external forcings have contributed to the observed warming hiatus, this study details the underlying experimental design, and the alternative forcings used to drive the Norwegian Earth System Model (NorESM) to bring them in line with each case realization of recent observations. These forcings are then used to drive two 30-member ensembles—a Reference ensemble using CMIP5-prescribed forcings and a Sensitivity ensemble using the alternative forcings discussed here. Section 2 outlines the broad experimental design and its underlying rationale. Section 3 describes the alternative forcings and compares them to the original forcings used. Section 4 describes the NorESM model and how the two 30-member ensembles were created. Section 5 outlines some principal characteristics and salient features of the ensembles. Section 6 provides a discussion of caveats, and section 7 summarizes with a hand-off to the second paper.
[Thorne et al., 2015] then compares the resulting ensembles to the available observational estimates of surface temperatures.

2. Experimental Design

As outlined in section 1, there exist multiple published hypotheses for the underlying causes of the recent warming hiatus that involve forcings and/or variability. The most enable tools to investigate these issues further are climate models, which can be run with different external forcings and be used to explore possible internal climate system variability as it is diagnosed by the model. But climate models are computationally expensive to run, making that careful forethought in experimental design is necessary and these are inevitable trade-offs required. Sexton et al. [2003] discussed this for atmosphere only runs, but to our knowledge no similar analysis has been applied to such an estimation approach in Earth System Models such as NorESM. Further, it is not clear how experimental design considerations depend upon the spatiotemporal scale of the question being assessed. Certainly, looking at decadal versus regional ESM models require very distinct modeling strategies. For the hiatus, our interest is in a relatively short period and understanding the regional signatures.

Our overriding concern was to attempt as comprehensively as possible to test the various existing hypotheses. These are not single right way to do this, especially given the available computer resources. The conundrum was to find a balance between exploring forcing uncertainty effects and internal variability effects in an ESM with no obvious a priori basis upon which to decide, in an informed manner, how to weight these two aspects. We know the models simulate the prescribed forcings as entirely deterministic. Given the notional computational resources available for 60 runs, we could run 60 runs each with some different forcings, two sets of 30 runs with only two different forcing scenarios, or anything between. The only way in which to explore forcing uncertainty is to run the models in multiple time-slice forcing scenarios. On the other hand, clearly assessing the role of variability alone requires consideration of a suite of runs with the same prescribed forcings, i.e., an ensemble.

Running 60 distinctly forced runs would make a clear distinction between internal variability and forced response possible, while running two 30-member ensembles yields only 2 degrees of freedom in the forced response dimension. Given no readily apparent decadal time scale internal ensemble spread in global mean surface temperatures in those CMIP5 submiss ions that undertook multiple ensemble runs and precursor analyses of e.g., Hawkins and Sutton [2009] showing the likely dominant role of internal variability in decadal predictability [see Thorne et al., 2015, Figure 1], we consciously chose to undertake two ensemble members of equal size. We reasoned that by requiring a far larger number of runs to elucidate variability on hiatus time scales than the effects of any distinction in forcings.

In summary, the use of two 30-member ensemble was felt at the outset of the project to be preferable in elucidating the potential roles of forcing and internal variability cleanly. Caveats and limitations associated with this choice, many of which naturally only become apparent at project completion, are retumed to in section 6.

3. Climate Forcings

This section describes which forcings were considered in creating a new set of ancillaries which takes advantage of the latest (at time of project inception in early 2014) available observational estimates of changes through at least 2012. Herein, the term "forcing" refers to "external forcing" as defined in the IPCC Assessment Reports [Planton, 2013], thus:

External forcing refers to a forcing agent outside the climate system causing a change in the climate system. Volcanic eruptions, solar variations and anthropogenic changes in the composition of the atmosphere and land use changes are external forcings. Orbital forcing is also an external forcing as the insolation changes with orbital parameters eccentricity, tilt and precession of the equinox.

In each case, the revised forcings are compared to those used in our 30-member Reference ensemble, which were identical to the forcings used in the version of NorESM extended historical runs submitted...
to the CMIP5 archive. For NorESM, the extended historical runs used RCP8.5 forcings post-2005 and the same choice is made in the Reference ensemble. The changes to the forcings include new measurements taken since 2005, as well as modifications made to the forcings prior to 2005. These modifications may come from inclusion of new data or due to changes in the interpretation of preexisting measurements.

As outlined in section 2, the forcings are treated entirely deterministically within the model. Therefore, by definition, there is absolutely no uncertainty in the applied forcings. Rather, the uncertainty arises in which of the forcing factors several possible forcing realizations to select and present to the model. Uncertainty in the selection of forcing histories to use and potential implications for the analysis in Part 2 is returned to in depth in section 6. For the remainder of this section, we concentrate upon describing the applied Sensitivity forcings, discussing where we sourced them to fulfill data provenance, justifying where forcings were not changed between the two ensembles, and finally, characterizing the differences in applied Radiative Forcing (RF) between the ensembles.

3.1. Greenhouse Gases

The primary greenhouse gases (GHGs) concentrations used in NorESM are carbon dioxide (CO₂), methane (CH₄), nitrous oxide (N₂O), CFC-11 (including additional CFCs and hydrochlorofluorocarbons (HCFCs) via nudging), and CFC-12. For the historical runs in the CMIP5 archive, observations of primary GHGs are used where available for the period of 1850 to 2005, although such observations only began during the last six decades and for several of these gases, reliable observations only started much more recently. The concentrations used in the Reference runs of this study are an average of the annual surface global mean mole fractions held by the National Oceanic and Atmospheric Administration (NOAA) and the Advanced Global Atmospheric Gases Experiment (AGAGE) [Prinn et al., 2000]. In the CMIP5 archive, the concentrations in the earlier part of the historical runs come from reconstructions based on air trapped in polar ice cores [Myhre et al., 2013]. For CO₂, for example, observations at Mauna Loa, Hawaii, and at the South Pole start in the late 1950s, with earlier values based on air extracted from ice cores and m. From 2005 until 2012, the concentrations in the Reference ensemble are taken from the Representative Concentration Pathways 8.5 scenario (RCP8.5). This is the same choice as in the submitted NorESM historical extended runs in the CMIP5 archive. For the Sensitivity runs, observations are used where available from 1980 until 2012 as detailed below.

3.1.1. Carbon Dioxide

The alternative carbon dioxide concentrations were obtained from the NOAA data set (NOAA.2014, http://www.esrl.noaa.gov/gmd/ccgg/trends/global.html.) and covered the period of 1980 to 2012. Figure 1 shows the absolute concentrations of CO₂ and resultant climate forcing relative to preindustrial forcing used in the Reference and Sensitivity ensembles. The main feature of the CO₂ concentrations is a steady increase between 1980 and 2012 that totals approximately 16% compared to the 1980 level or an increase in forcing of 0.81 W m⁻². The concentrations are alm. identical between the Reference and Sensitivity ensembles, with only a small difference seen after 2005. The largest difference is seen in 2012 and equates to a decrease of 0.02 W m⁻².

3.1.2. Methane

The methane concentrations for the alternative forcings were obtained from the World Meteorological Organization (WMO) data set [Tutina et al., 2009]. This only covered the period of 1984 until 2012, so concentrations between 1980 and 1983 were maintained at their CMIP5 values (Figure 1). There is a marked and consistent difference between the CMIP5 concentrations used in the Reference runs and the alternative concentrations used in the Sensitivity runs. This results in an increase of around 2.5% in the forcing relative to preindustrial caused by methane over 0.01 W m⁻². The offset is relatively invariant with time after the initial shock in 1983/1984. This is in the spin-up period for the ensemble (section 4) and predatesthe hiatus, which is the focus of the present study, by approximately 15 years.

3.1.3. Nitrous Oxide

The nitrous oxide, the alternative concentrations were also obtained from the WMO data set [Tutina et al., 2009]. Since N₂O has been systematically observed since the late 1970s, these cover the full period of the ensemble runs from 1980 until 2012. The alternative forcings show a mean difference over the hiatus period of 0.0003 W m⁻² from the original concentrations used in CMIP5 (Figure 1). Nitrous oxide concentrations have increased steadily over this period resulting in an increase in forcing of 0.07 W m⁻².
3.1.4 CFC-11

The most complete record of CFC-11 concentration was found in the AGAGE data set NASA-AGAGE, 2014, http://agage.eas.gatech.edu/data.htm, and this was used to replace the concentrations of CFC-11 for the Sensitivity runs. It should be noted that in NorESM, the CFC-11 e is adjusted to incorporate the effect of trace gases including long-lived GHGs and ozone-depleting substances. Figure 1 compares the adjusted CFC-11 concentrations and resultant climate forcing in the Reference and Sensitivity forcings. The largest changes occur from 2009 onward and result in an increase in forcing relative to preindustrial levels of approximately 0.004 W m$^{-2}$. The concentrations of the trace gases have been converted to CFC-11 equivalent concentrations using the relative radiative efficiencies of the individual trace gases. The right axis in each plot shows the forcing relative to preindustrial levels. Despite appearances of linearity over such a small range, the right axes for CO$_2$, CH$_4$, and N$_2$O are nonlinear.

![Figure 1](image_url)

**Figure 1.** (top left) Concentrations of carbon dioxide, (top right) methane, (middle left) nitrous oxide, (middle right) CFC-11 adjusted, and (bottom right) CFC-12 in NorESM for the Reference (blue) and Sensitivity (red) ensembles. The separation of (bottom left) CFC-11 adjusted into its components of CFC-11 (black solid) and trace gases (black dashed) is also given. The concentrations of the trace gases have been converted to CFC-11 equivalent concentrations using the relative radiative efficiencies of the individual trace gases. The right axis in each plot shows the forcing relative to preindustrial levels. Despite appearances of linearity over such a small range, the right axes for CO$_2$, CH$_4$, and N$_2$O are nonlinear.

3.1.4 CFC-11

The most complete record of CFC-11 concentration was found in the AGAGE data set NASA-AGAGE, 2014, http://agage.eas.gatech.edu/data.htm, and this was used to replace the concentrations of CFC-11 for the Sensitivity runs. It should be noted that in NorESM, the CFC-11 e is adjusted to incorporate the effect of trace gases including long-lived GHGs and ozone-depleting substances. Figure 1 compares the adjusted CFC-11 concentrations and resultant climate forcing in the Reference and Sensitivity forcings. The largest changes occur from 2009 onward and result in an increase in forcing relative to preindustrial levels of approximately 0.004 W m$^{-2}$. To incorporate the radiative effect of the trace gases, they are converted to an equivalent concentration of CFC-11 and added to that e, giving rise to the adjusted CFC-11. This is done using the following equation, which utilizes the ratio of radiative efficiencies of the trace gases compared to that of CFC-11.

\[
\text{CFC-11 equivalent of } \frac{\text{X ppt}}{\text{CFC-11}} = \text{CFC-11} \times \frac{\text{radiative efficiency of X}}{\text{radiative efficiency of CFC-11}}
\]

NorESM includes forcings from numerous trace gases in this way, as shown in Table 1. Figure 1 shows the adjusted CFC-11 along with the true measured CFC-11 and the equivalent concentration of all of the summing contributory trace gases combined. Until the late 1980s, the combined contribution of the trace gases...
was comparable to that of CFC-11; however, the success of the Montreal Protocol [UNEP Ozone Secretariat, 1987] and its subsequent amendments means that CFC-11 concentrations have been steadily decreasing since around 1990, while the trace gases have steadily increased. The net effect is that since 2008, the combined forcing of the trace gases is more than double that of CFC-11 alone.

Since the trace gases are generally not as well monitored as GHGs, it was not possible to use alternative concentrations for any of them, especially extending as far back as 1980. Their concentrations were therefore replaced using whatever observations could be found. CF$_4$ and C$_2$F$_6$ were both replaced for 2004 until 2012 from the AGAGE data set (NASA-AGAGE, 2014, http://agage.eas.gatech.edu/data.htm). CFC-113 was also taken from the AGAGE data set (NASA-AGAGE, 2014, http://agage.eas.gatech.edu/data.htm.) for the period of 1985 until 2012. NOAA's Halocarbons and other Atmospheric Trace Species Group (HATS) provided the alternative concentrations of HFC-23, HFC-32, HFC-43-10, HFC-125, HFC-134a, HCFC-141, CH$_3$Br, CH$_3$Cl, SF$_6$, HCFC-22, HCFC-142, Halon-1211, and Halon-1301 from 1994, 1992, 1993, 1995, 1991, 1992, and 2004 respectively (NOAA HATS, 2014, http://www.esrl.noaa.gov/gmd/hats/data.html). Finally, CFC-114 and CFC-115 were taken from the AGAGE data set (NASA-AGAGE, 2014, http://agage.eas.gatech.edu/data.htm), starting from 2006. This means that 27% of the annual concentrations of the trace gases were replaced during the period of 1980 until 2012. This increases to around 40% for the period of 1990 until 2012 and to around 70% for 2000 until 2012. The net change in forcing by these trace gases over the hiatus period is 0.003 W m$^{-2}$.

### 3.1.5. CFC-12

The AGAGE data set provided the alternative CFC-12 concentrations (NASA-AGAGE, 2014, http://agage.eas.gatech.edu/data.htm). These cover the full period of 1980 until 2012 and are shown in Figure 1 along with the CFC-12 concentrations from the Reference ensemble. The alternative forcings relative to preindustrial levels are consistently higher than those found in CMIP5 runs, with an even difference of around 12 ppt, or 0.004 W m$^{-2}$.

### 3.1.6. Summary of GHGs

Most of the GHGs show little change from the CMIP5 runs used in the Reference runs, with the largest change arising from the alternative methane concentrations. The combined effect of all of the GHGs is shown in Figure 2. The average difference between these forcings in the two ensembles relative to preindustrial levels is ~0.003 W m$^{-2}$.

### Table 1. Long-Lived Trace Gases and Ozone-Depleting Substances Included in NoR-ESM as Part of the Adjusted CFC-11 Concentration

<table>
<thead>
<tr>
<th>Trace Gas</th>
<th>Radiative EF (ppt)</th>
<th>Mean Concentration Between 1980 and 2012 (ppt)</th>
<th>Equivalent Concentration of CFC-11 (ppt)</th>
<th>Modified From</th>
</tr>
</thead>
<tbody>
<tr>
<td>CFC-11</td>
<td>0.25</td>
<td>239</td>
<td>239</td>
<td>1980</td>
</tr>
<tr>
<td>C$_2$F$_6$</td>
<td>0.1</td>
<td>71.2</td>
<td>28.5</td>
<td>2004</td>
</tr>
<tr>
<td>C$_2$F$_4$</td>
<td>0.26</td>
<td>2.6</td>
<td>2.7</td>
<td>2004</td>
</tr>
<tr>
<td>C$_2$F$_4$I$_4$</td>
<td>0.04</td>
<td>0.62</td>
<td>0.03</td>
<td>–</td>
</tr>
<tr>
<td>HFC-23</td>
<td>0.39</td>
<td>12.6</td>
<td>9.6</td>
<td>–</td>
</tr>
<tr>
<td>HFC-32</td>
<td>0.11</td>
<td>1.3</td>
<td>0.6</td>
<td>–</td>
</tr>
<tr>
<td>HFC-43-10</td>
<td>0.4</td>
<td>0.1</td>
<td>0.2</td>
<td>–</td>
</tr>
<tr>
<td>HFC-125</td>
<td>0.23</td>
<td>2.0</td>
<td>1.8</td>
<td>–</td>
</tr>
<tr>
<td>HFC-134a</td>
<td>0.16</td>
<td>172</td>
<td>110</td>
<td>1994</td>
</tr>
<tr>
<td>HFC-143a</td>
<td>0.13</td>
<td>33</td>
<td>1.7</td>
<td>–</td>
</tr>
<tr>
<td>HFC-227ea</td>
<td>0.26</td>
<td>0.3</td>
<td>0.4</td>
<td>–</td>
</tr>
<tr>
<td>HFC-245fa</td>
<td>0.28</td>
<td>2.0</td>
<td>2.3</td>
<td>–</td>
</tr>
<tr>
<td>SF$_6$</td>
<td>0.52</td>
<td>3.8</td>
<td>8.0</td>
<td>1995</td>
</tr>
<tr>
<td>CFC-113</td>
<td>0.3</td>
<td>66.7</td>
<td>80.0</td>
<td>1985</td>
</tr>
<tr>
<td>CFC-114</td>
<td>0.31</td>
<td>153</td>
<td>189</td>
<td>2006</td>
</tr>
<tr>
<td>CFC-115</td>
<td>0.18</td>
<td>6.4</td>
<td>4.6</td>
<td>2006</td>
</tr>
<tr>
<td>CFC-117</td>
<td>0.13</td>
<td>975</td>
<td>507</td>
<td>–</td>
</tr>
<tr>
<td>MCF</td>
<td>0.06</td>
<td>73.2</td>
<td>176</td>
<td>–</td>
</tr>
<tr>
<td>HCFC-22</td>
<td>0.2</td>
<td>124.9</td>
<td>99.9</td>
<td>1991</td>
</tr>
<tr>
<td>HCFC-134a</td>
<td>0.14</td>
<td>8.9</td>
<td>5.0</td>
<td>1992</td>
</tr>
<tr>
<td>HCFC-142</td>
<td>0.2</td>
<td>8.7</td>
<td>7.0</td>
<td>1992</td>
</tr>
<tr>
<td>Halon-1211</td>
<td>0.03</td>
<td>3.0</td>
<td>3.6</td>
<td>1992</td>
</tr>
<tr>
<td>Halon-1301</td>
<td>0.02</td>
<td>2.0</td>
<td>2.6</td>
<td>2004</td>
</tr>
<tr>
<td>Halon-2402</td>
<td>0.33</td>
<td>0.3</td>
<td>0.4</td>
<td>2004</td>
</tr>
<tr>
<td>CH$_3$Br</td>
<td>0.01</td>
<td>8.7</td>
<td>0.3</td>
<td>1993</td>
</tr>
<tr>
<td>CH$_3$Cl</td>
<td>0.01</td>
<td>520.6</td>
<td>20.8</td>
<td>1999</td>
</tr>
</tbody>
</table>
over the period of 1980 to 2012 is 0.014 W m⁻², approximately 0.01 W m⁻² of which comes from the alternative methane, while around 0.004 W m⁻² of which comes from the alternative CFC-12. However, the combined effect of the alternative greenhouse gas forcings, including the long-lived GHGs and ozone-depleting substances, is an increase of only around 0.6% of total GHG forcing relative to preindustrial and corresponds to a relative error of around 1.2% on the change of forcings between 1980 and 2012. The mean net change in forcing from GHGs including long-lived trace gases and ozone-depleting substances for the hiatus period is 0.012 W m⁻².

3.2. Anthropogenic Aerosols

NorESM includes three types of aerosols other than volcanic: black carbon (BC), sulfur dioxide (SO₂), and primary organic matter (POM) [Kirkevåg et al., 2013; Iversen et al., 2013]. These are read into the models as emissions (gm⁻²s⁻¹) from both fossil fuel and biomass burning sources. Black carbon is also read in based on emissions from air traffic. These are defined at every location and on every level as monthly averages; however, they are only read in for one single annual cycle once per decade. For example, in the historical runs in CMIP5, 12 monthly averages were read in for 1850, 1860, etc. until 1990 and 2000, with a single annual cycle included for 2005, the end of the simulation. However, in the Reference runs discussed in this work, the 2005 point is ignored and interpolation is performed between the 2000 historical emissions and the 2010 RCP8.5 emissions. NorESM performs a linear weighting to determine aerosol emissions for the years between the input decades.

Alternative aerosol emissions were obtained from the Norwegian Meteorological Institute and were taken from the EU-Eclipse V4a data set [Klimont et al., 2013a, 2013b]. These included new annual cycles for each decade between 1970 and 2000, and for each year between 2005 and 2010. They also included a single annual cycle and for 2020 to ensure that NorESM could interpolate emissions for 2011 and 2012. The alternative emissions do not include new values for black carbon emitted by air traffic; these data were unavailable.

The original Reference and new Sensitivity emissions are shown in Figure 3. In order to ascertain the radiative impact of the perturbations to the aerosol emissions, we perform a set of auxiliary simulations. Because the aim is to estimate the effect relative to preindustrial (PI) atmospheric burdens, these sets of runs are required. These runs are undertaken utilizing, respectively, the following: preindustrial (mainly natural) emissions with no time variation beyond a seasonal cycle (PI), the varying CMIP5 emissions used in Reference (Ref), and the varying Eclipse V4a emissions used in Sensitivity (Sens). In all cases, the radiation (direct and cloud nucleation) and cloud feedbacks (not indirect effects) are computed in "online" mode using the prescribed aerosol concentrations provided by the National Center for Atmospheric Research as part of the Community Atmosphere Model. The aerosol-climate calculations that result from the distinct emission scenarios in PI, Ref, and Sens. [Kirkevåg et al., 2013]. The nalaerosol climatic forcing was estimated as the difference between Ref/PI and Sens/PI in the calculated top of atmosphere (TOA) net radiative forcing.

The alternative emissions show higher levels of black carbon after 2000, which provide positive climate forcing by decreasing the albedo and allowing the atmosphere to absorb more heat. They also show higher levels of SO₂, which serve to strongly cool the atmosphere. Finally, primary organic matter (POM) emissions...
are lower with a marked relative minimum in 2009. POM emissions are “brown” and so have both positive and negative direct effects. However, their indirect effects are undoubtedly negative (RF [Myhre et al., 2013, Table 8.4]). By design, the Sensitivity aerosol emissions exhibit substantially greater interannual variability than those used in Reference over the hiatus. The differences between the aerosol emissions are substantial over this period. All of the aerosol undoubtedly contribute to the difference in RF in Figure 3 (bottom right). However, Figure 3 suggests that the changes in POM are dominant as the peak in 2009 coincides with a relative minimum in BC and the strong minimum in POM. Maximum relative changes note that Figure 3’s y-axis range do not start from zero are greater in POM (~25%) than in SO₂ (~5%) or BC (~10%). If BC were the dominant factor, the peak in the difference between Reference and Sensitivity RF would occur in 2010. Further analysis beyond the current study would be needed to diagnose the impacts of the individual aerosols in the model.

Regardless, the mean net difference in radiative forcing between the Reference and the Sensitivity aerosol over the hiatus period is 0.11 W m⁻², with the largest difference occurring in 2009 and reaching 0.28 W m⁻². Changes in all three sets of aerosol emissions are likely to be important, but the change in POM emissions dominates the change in overall RF between the Reference and the Sensitivity forcings.

3.3. Solar

Solar forcing is included in NorESM as an annual mean value of total solar insolation, with diurnal and seasonal cycles in solar forcing applied through the model code. The values used in the NorESM CMIP5 simulations were based on Lean [2000], and they were adjusted by a factor of 0.9965 based on the work of Wang et al. [2005] to bring them in line with observations from the Total Irradiance Monitor (TIM), a space-based instrument launched in 2003 as part of NASA’s Earth Observing System Solar Radiation and Climate Experiment (SORCE). From 2009 onward, the solar forcing in CMIP5 was a repeat of the last four cycles of the historical
Some concerns have been raised over this approach since the cycle from 1996 until 2008 was in reality only 12.2 years long, not 13 years as is in the model [Lean and Rind, 2009]. Furthermore, the two cycles prior to 1996 were also atypical, being two of the shortest on record. The solar forcing in our 30-member ensemble of Reference runs was identical to those of the CMIP5 simulations, including the repeated cycle from 2009 onward.

The Sensitivity ensemble uses the solar forcings from the Physikalisch-Meteorologisches Observatorium Davos (PMOD) that are based on a combination of data from several instruments:

- Active Cavity Radiometer Irradiance Monitor-I (ACRIM-I)
- ACRIM-II
- ACRIM-III
- Variability of Solar Irradiance and Gravity Oscillations (VIRGO)
- Total Irradiance Monitor (TIM)
- Earth Radiation Budget Experiment (ERBE).

The PMOD data are generally considered the most accurate set of solar forcings, given how it was constructed [Frohlich and Lean, 1998; Frohlich, 2006]. Uncertainty in solar forcings is discussed further in section 6.1.1.

It should be noted that while the difference between the PMOD solar forcing used in Sensitivity and that of the original NorESM CMIP runs used in Reference reaches 0.5 W m⁻² in 2009, this is the total solar irradiance. The Earth receives this radiation effectively as a two-dimensional disk rather than a three-dimensional sphere, and hence with a quarter of the surface area that it truly has. Therefore, the effective forcing is shown on the left axis in Figure 4, and this is one quarter of the actual TOA forcing. The difference between the PMOD forcings and the NorESM original forcing is therefore up to around 0.125 W m⁻², although this varies with time, with the mean over the hiatus period being 0.08 W m⁻². This is an order of magnitude greater than the change in GHGs as discussed in section 3.1 (Figure 2) and potentially comparable, but of opposite sign, to the highly uncertain effect of aerosol changes (section 3.2).

### 3.4. Volcanic

The forcing effect of volcanic aerosols is included in NorESM through a zonal mean long ratio specified for each latitude and on each vertical level. In practice, the model uses a zonal mean column mass of aerosols distributed over the vertical levels by a simple shape function. The column mass of volcanic aerosols used in NorESM was based on the work of Sato et al. [1993]. The last volcanic eruption included in this data set is Pinatubo in 1991. Traces of this eruption exist until 2001, after which there is no volcanic aerosol forcing in NorESM. Since Pinatubo, there have been a series of eruptions each adding small quantities of aerosols to the atmosphere [Santer et al., 2014; Ridley et al., 2014].

Alternative volcanic aerosol concentrations were obtained from the NASA Goddard Institute for Space Science (NASA-GISS, 2014, http://data.giss.nasa.gov/modelforce/statens/) and are based on the update of the tabulation of stratospheric aerosol optical thickness given in Sato et al. [1993]. Figure 5 compares the column mass of volcanic aerosols for the original NorESM concentrations and the alternative concentrations, along with the associated latitudinally integrated forcings. The forcings were calculated using the same formula as was used in IPCC AR5 (Table 8.SM.9), where RF = 25 * [Amospheric Optical Depth in W m⁻²] [Hansen et al., 2005]. The weaker volcanoes are clearly missing from the Reference concentrations. Furthermore, the aerosols from the Pinatubo eruption have decreased in peak magnitude in the alternative concentrations but have a greater presence at middle and high latitudes. Assuming the presence of a long-term multidecadal response to episodic large volcanoes [Santer, 2010], such a missed eruption may have a significant influence on the behavior in the hiatus period even though such an eruption occurred prior to this. Specifically, the aerosol concentrations in the Pinatubo eruption have decreased in peak magnitude in the alternative concentrations, but have a greater presence at middle and high latitudes. Assuming the presence of a long-term multidecadal response to episodic large volcanoes [Santer, 2010], such an eruption may have a significant influence on the behavior in the hiatus period even though such an eruption occurred prior to this. Specifically, the aerosol concentrations in the Pinatubo eruption have decreased in peak magnitude in the alternative concentrations. This complicates a clean comparison of the likely effect of the forcing differences.
Themea n change in forcings for 1980 to 2012 between the Reference and Sensitivity forcings is 0.17 W m$^{-2}$. Over the hiatus period, the alternative forcings exhibit systematically greater loading that would impart a cooling influence. The mean change in forcings over the hiatus period is 0.08 W m$^{-2}$, which compares well to the estimate from Box 9.2 of the IPCC report of around 0.09 W m$^{-2}$.

3.5. Forcings Not Altered

3.5.1. Stratospheric Water Vapor

Between 2000 and 2005, the stratospheric water vapor shows a sharp decline; however, the concentrations have since risen steadily until 2012 [Dessler et al., 2013]. A recent study by Urban et al. [2014] has shown that another sharp drop in stratospheric water vapor concentrations has occurred over the last 18 months. Decreases in stratospheric water vapor are expected to warm the stratosphere but cool the troposphere [Solomon et al., 2010]. Although this idea was disputed by Kaufman et al. [2011], who found no statistically significant relationship between changing stratospheric water vapor and surface temperatures, Dessler et al. [2013] quantified the feedback of stratospheric water vapor to be approximately 0.3 W m$^{-2}$K$^{-1}$.

Unfortunately, the stratospheric water vapor concentrations are not read into NorESM and cannot therefore be changed directly. This, to some extent, re-acts the reality that stratospheric water vapor is definitely radiatively important, but not necessarily in a feedback role. In current climate models, the effect of stratospheric water vapor is typically included to account for radiative effects. If we change the water vapor on those levels for radiative purposes, we have to include the effect of stratospheric water vapor, the effect of which is to change the longwave and shortwave components of the code. This would have the effect of changing the water vapor on those levels for radiative purposes. However, the resulting temperature response to this change was contrary to existing theory and literature on this subject, and no explanation could be determined for this.

Since we currently have no other method for adjusting the stratospheric water vapor in the NorESM model, and the only method we do have produces results that are inconsistent with the current literature, stratospheric water vapor was not altered in the Sensitivity runs in this study.

3.5.2. Ozone

NorESM includes ozone concentrations as a forcing ancillary field. These concentrations are spatially complete, being read in for every grid point and every model level. However, the ozone concentrations are included in the model on a temporally intermittent basis. For example, in the historical model runs, the monthly ozone concentrations for a single year are read into the model every 10 years from 1850 until 1990, and every 5 years from 1995 until 2005 (Figure 6).

Figure 5. Column mass of volcanic aerosols in kg m$^{-2}$ by latitude with the meridionally integrated forcings underneath, for the (top) Reference forcings from the NorESM CMIP5 runs and the (bottom) Sensitivity forcings. These are both based on the work of Sato et al.

Figure 6. Monthly ozone concentrations input to NorESM once every 10 years until 1990 and every 5 years until 2005.
The input frequency of ozone into NorESM is insufficient to resolve its decadal scale impacts which would require forcing estimates that varied continuously in space and time. While complete observations of vertically resolved ozone across the globe for the whole period of interest are available, since no hypothesis has been proposed in the literature professing the role of ozone in causing the currently observed warming hiatus, the ozone driving field has remained unchanged in our current experiment. A more complete study would be able to incorporate more accurate ozone fields into NorESM. This would not only allow for the investigation of the impact of changing ozone concentration on decadal timescales but would also help quantify the importance of ozone on a temporal ozone representation in future runs.

3.5.3. Land Surface Changes

NorESM includes the Community Land Surface Model version 4 as its land surface component. The land use data was unchanged between the Reference and Sensitivity runs and is the same as the land use data specified for CMIP5 simulations. These consist of land use estimates from the History Database of the Global Environment, smoothed to combine seamlessly with the future predictions under the RCP8.5 scenario. The land use data was not modified in this study partly because at the time of the project execution, land use changes had not been posited as a major factor in causing or explaining the global warming hiatus. Furthermore, we did not find an observational dataset of land use that was significantly different from the standard input of NorESM.

3.6. Summary of Combined Radiative Effects of Changed Forcings

In this section, we have documented changes to a number of forcing ancillaries acting on the observational understanding at the time of project inception in early 2014. Changes were made to GHGs and long-lived greenhouse gases (LLGHGs) (section 3.1), anthropogenic aerosols (section 3.2), solar radiation (section 3.3), and volcanic aerosols (section 3.4). All of the forcings that were revisited were added with the exception of solar forcing, which was not modified. In this study, we examined the sensitivity of realizing the role of ozone in causing or explaining the global warming hiatus. Put them one, we did not find an observational dataset of land use that was significantly different from the standard input of NorESM.

4. Model Configuration and Ensemble Creation

4.1. Model System

This study uses the medium resolution configuration of the Norwegian Earth System Model version 1 (NorESM 1-M) which has provided output to the fifth Coupled Model Intercomparison Project (CMIP5)
The model is based on the Community Climate System Model version 4 (CCSM4) [Gent et al., 2011] with important changes being the replacement of the z-level ocean component with an isopycnic coordinate ocean model, in providing the representation of water masses, and the employment of advanced chemistry-aerosol-cloud-radiation interaction schemes, enabling the model to account for the indirect aerosol feedback. The sea ice and land components—the Los Alamos Sea Ice Model and the Community Land Model—have been adopted from CCSM4 without changes. The atmosphere and land components are conserved on a regular horizontal grid with a 1.9°×2.5° resolution. In the vertical, the atmosphere component comprises 26 hybrid-σ levels extending up to 3 hPa. The ocean and sea ice components are conserved on a curvilinear horizontal grid with 1° resolution along the equator and the northern grid singularity shifted over Greenland. In the vertical, the ocean component comprises a two-layer bulk mixed-layer representation with 51 isopycnic layers below. A detailed description of the model and evaluation of standard simulations are given in Bentzen et al. [2013] and Iversen et al. [2013].

4.2. Experimental Setup

We have performed two sets of simulations, one Reference ensemble which uses the default CMIP5 forcings, and one Sensitivity ensemble which uses the alternative forcings outlined in section 3. The simulations covered the period of 1980 until 2012. This period was chosen for these reasons.

1. It includes the period of interest for studying the observed hiatus of global warming.
2. Since the hiatus started in the 1990s, this choice allows the model to evolve a decade in order to adjust to the observed forcings, thus avoiding sharp changes in the forcings during the period of interest.
3. Starting over a decade prior to the period of interest allows the individual runs to spin-up and diverge such that they exhibit a broad range of basic states of the modeled Earth system (including cryosphere and ocean) consistent with the model physics by the start of the period of study. The end date is driven by the availability of some of the Sensitivity runs which use forcings being solely through 2012.

To allow for a robust detection of any differences in forced response and an elucidation of the effects of model internal variability on decadal time scales, each ensemble comprises a total of 30 simulations. The initial conditions are generated from the three ensemble members of NorESM’s CMIP5 historical experiment [Bentsen et al., 2013] as follows: Both sets are split into two subsets of 10 simulations each. The simulations of each subset are initialized with the 1980-01-01 state of the same CMIP5 NorESM historical simulation, i.e., the simulations of subset 1 are initialized with the state of the first CMIP5 historical realization, the simulations of subset 2 are initialized with the state of the second CMIP5 historical realization, and the simulations of subset 3 are initialized with the state of the third CMIP5 historical realization. The three historical runs that were used each run off the model control run separated by several decades and by 1980 have each been running for 130 years. Within each subset, the initial spread is generated by adding microscopic (0.1 °C) noise to the ocean mixed-layer temperatures. The internal spread then grows with time and at least for the atmosphere reaches saturation after a decade or so, i.e., before our period of primary interest.

5. Characterization of the Model

As discussed previously, various alternative hypotheses other than forcings involving internal variability or feedbacks within the system have been proposed to explain the observed hiatus in global warming. Before detailed analysis of the two ensembles can be undertaken, it is important to place NorESM in context of the wide range of CMIP5 models and to establish whether or not the introduction of the alternative forcings has in fact changed the alternative mechanisms in the model. Figure 8 shows the Equilibrium Climate Sensitivity (ECS) against the Transient Climate Response (TCR) for a range of CMIP5 models (this is a modified version of Figure 9.42 from Chapter 9 of IPCC AR5) [Flato et al., 2013]. NorESM has an ECS of approximately 2.8°C and a TCR of approximately 1.4°C and thus lies well within the spread of CMIP5 models. While it does lie further from the CMIP5 mean than the CCSM4 mean used in the work of Santer et al. [2014], NorESM is apparently not unusual among the CMIP5 models for its response to changes in climate forcings. The power spectrum density of global mean temperature variance in the historical simulations of CMIP5 models is given in Figure 9 (this is a modified version of Figure 9.33 from Chapter 9 of IPCC AR5) [Flato et al., 2013]. NorESM is generally consistent with observations and shows reasonable power at hiatus-like time scales that is in broad agreement with other model assessments and observations.
5.1. Changes in Ocean Heat Content

Various studies have discussed the possibility of increased heat sequestration into the world oceans [e.g., Mehle et al., 2011, 2013, 2014; Chen and Tung, 2014; Allan et al., 2014; England et al., 2014]. Figure 10 shows the anomalies of global OHC for the two ensembles in the shallow ocean and the deep ocean, defined as the surface to 325 m and below 325 m, respectively. Both ensembles show a similar increase occurring between 1980 and 2012 of approximately 100 ZJ in the shallow ocean and 200 ZJ occurring in the deeper ocean. The total OHC increase of approximately 300 ZJ is similar to the observational estimate reported in the IPCC report [Rhein et al., 2013]. Both ensembles show a similar magnitude local minima around 1992 resulting from the Pinatubo eruption that are greatest in the uppermost 325 m.

In the shallow ocean, the Sensitivity ensemble shows increased variability over the hiatus period, especially after 2006 when the RCP8.5 forcings were replaced with observational estimates. The Sensitivity ensemble in the deep ocean has approximately the same variability as the Reference ensemble but with a slightly decreased mean. This is to be expected since the total ocean heat content should be very similar to the total Earth system heat content which is controlled by the forcings. Since in the net the Sensitivity forcings show a decrease, we would expect the total OHC anomaly to also decrease somewhat. Figure 10 suggests that the trends in OHC during the hiatus period of 1998 to 2012 are slightly higher in the Sensitivity ensemble than in the Reference ensemble for both the shallow and deep oceans. Figure 10 further suggests that there is a greater increase in the deep ocean and least heat content increase in the shallow ocean during the hiatus period than in the warming period of the early 1990s. To elucidate this, a comparison of the trends in OHC for deep and shallow oceans over the hiatus period is given for all 30 members of both the Reference and Sensitivity ensembles (Figure 11). The trends were calculated from the globally averaged OHC annual series by ordinary least squares (OLS) regression over the period 1998–2012. As expected [Meehl et al., 2013], there is an anticorrelation whereby when near-surface uptake is relatively rapid, deep ocean uptake is relatively slow and vice versa.

The trends of the total OHC were examined for the hiatus period and found to be 13.4 ZJyr$^{-1}$ and 13.9 ZJyr$^{-1}$ for the Reference and Sensitivity ensembles, respectively. A student’s t test was applied to show that the trends in OHC between the two ensembles were different at the 5% significance level. Overall, the inclusion of the alternative forcings appears to have had a minimal impact on the behavior of the modeled OHC in the shallow ocean.

In the deep ocean, the alternative forcings reduce the mean OHC in the Sensitivity ensemble compared to the Reference ensemble, but increased trends over the last 15 years act to decrease this difference by the end of the run in 2012. A recent paper by Morii et al. suggested the need for a higher number of ensemble members (at least 80) in order to detect a significant response in surface air temperature to changes in...
boundary conditions. As will be discussed in Part 2, some runs in the Sensitivity ensemble began the hiatus period with a La Niña situation and end with an El Niño situation than in the Reference ensemble. This could in part explain the difference in OHC observed here.

5.2. ENSO Variability and Trends

The ocean temperatures of the Eastern Pacific have been of particular interest due to their role in driving the El-Niño–Southern Oscillation (ENSO). Comparing the temperature anomalies over the ENSO 3.4 region shows that the frequency, magnitude, and seasonality of the variations are all comparable to those in the observations for both the Reference and Sensitivity ensembles (Figure 11). The periodogram highlights that a broad peak in variability occurs between 2 and 7 years, in agreement with the observed variability of ENSO. The introduction of the Sensitivity forcings appears to have had no significant impact on ENSO as it is reproduced by NorESM, based on a student's test of the power spectrum.

5.3. TOA Radiation

The net imbalance of radiation at the top of atmosphere provides insight into both the changes in radiative forcing and the changes in climate response. This net imbalance has been shown to capture interannual variability in the climate system from both volcanoes and ENSO [Allan et al., 2014] and has been shown to change in response to lower frequency unforced modes such as the Interdecadal Pacific Oscillation [Brown et al., 2014]. Figure 12 shows this net imbalance of TOA radiation from the Reference and Sensitivity ensembles, compared to the observations from the CERES and the Earth Radiation Budget Satellite (ERBS) as compiled by Allan et al. [2014]. The observations only begin in 1985 at the start of the ERBS data. The volcanic eruptions of El Chichón and Pinatubo are clearly visible in the 1980s and 1990s, respectively, although the observations do not cover the period of the El Chichón eruption. There is reasonable concurrence between the observational record and the model ensemble over the period that they overlap.
5.4. Arctic Sea Ice Extent

Finally, the Arctic sea ice extent, previously linked to cooling over midlatitude Eurasia [Petoukhov and Semenov, 2010; Outten and Esau, 2012; Honda et al., 2009], is shown in Figure 14 for the two ensembles and compared to observational estimates from the National Snow and Ice Data Centre [Fetterer et al., 2002]. The NorESM model has clear deficiencies in its seasonal cycle of Arctic region sea ice, having a seasonal cycle that is too small, even though the mean extent over the calendar year may be broadly correct. The wintertime maximum is too small, and the summertime minimum too large. This misrepresentation may limit the value of NorESM for considering at least some aspects of sea ice-atmosphere feedbacks.

Differences in the trends and variability of sea ice extent between the two ensembles are not significant at the 5% level according to a student’s t-test; hence, the introduction of the Sensitivity forcings has not significantly altered the potential source of climate feedback. In both ensembles, the model shows a slightly slower decrease in the wintertime sea ice extent than is seen in observations, and it fails to reproduce the steep drop in summertime sea ice observed over the past 10–15 years.

Figure 12. Temperature anomalies in the top) ENSO 3.4 region from HadCRUT4 (black) [Morice et al., 2012], the Reference ensemble (blue), and Sensitivity ensemble (red), with the anomaly varying (bottom) periodogram. The solid vertical lines on the periodogram indicate periods of 2 years and 7 years.

Figure 13. Net imbalance in radiation at the top of the atmosphere from the Reference ensemble (blue) and Sensitivity ensemble (red), compared to observations from CERES and ERBS [Allen et al., 2014]. Plotted are the 95% bootstrap confidence bounds (shading), together with ensemble mean (solid line).
6. Discussion

The recent IPCC report [Flato et al., 2013] proposed that the reduced trend in the external forcings, largely as a result of solar and volcanic factors with a poorly quantified contribution from anthropogenic aerosols, was a factor roughly equal in importance to internal variability in explaining the currently observed warming hiatus. To investigate the role of forcings, the forcings in the Norwegian Earth System Model due to primary greenhouse gases, solar irradiance, volcanic aerosols, anthropogenic and biomass burning aerosols, and a number of long-lived GHGs and ozone-depleting substances have been modified based on multiple state-of-the-art observational sources. These modifications not only included new values for the period of 2006 to 2012, i.e., beyond the CMIP5 historical run, but also modifications for the period prior to 2006 based on improved understanding and associated data set innovations and updates. To assess the possible role of internal variability, each set of forcings was run 30 times starting from distinct initial conditions with 18 years of spin-up to ensure divergence. There are a number of caveats required in both the forcings and ensemble creation choices which may impact the subsequent analysis and that we outline in this section.

6.1. Forcings

The alternative forcings discussed in section 3 show differences between the Reference and Sensitivity forcings; however, the new forcings show only a small change (relative to the common signal of increasing net forcing) from those applied in the Reference runs, which were the CMIP5 forcings extended with RCP8.5. For the hiatus period of 1998–2012, the Sensitivity forcings showed differences from the Reference forcings of approximately 0.02 W m\(^{-2}\) due to the GHGs and long-lived trace gases, 0.08 W m\(^{-2}\) due to decreased solar forcing, 0.11 W m\(^{-2}\) due to anthropogenic aerosol changes, and 0.08 W m\(^{-2}\) due to increased volcanic activity. Thus, the largest differences were found in the solar irradiance, volcanic, and tropospheric aerosols.

There is an implicit assumption that the data sets used to create the Sensitivity forcings in this work were not themselves extended or extended unless the producer of that data set believed it was an improvement not to do so. While this is self-evident for data sets that have solely been extended with new observations that did not previously exist, a question could be raised about the improved extended data obtained by modifying preexisting data (e.g., volcanic aerosols from Pinatubo). For this reason, the authors make no claim in this work that the Sensitivity forcings are better than those in the Reference runs, only that they have been modified to bring them in line with current understanding and are different from those in the Reference runs. Further, it should be noted that the black carbon emissions from aircraft, stratospheric water vapor, and sea ice change, and ozone concentrations were not modified due to either a lack of suitable observational data or a lack of a suitable method for modifying the forcing in NorESM. The implications of these emissions by de Ruijter and de Milt are unknown. However, based upon Myhre et al. [2013], the forcing effects of ozone or black carbon from aircraft case expected to be small compared with the impact from other aerosols and the solar forcing.

As detailed in section 2, there was an inevitable trade-off in the expert ental design whereby we allowed only 2 degrees of freedom in the forcings. Yet it is beyond dispute that there is uncertainty in many of the applied forcings. Those forcings which changed most substantially between Reference and Sensitivity,
solar, volcanic, and anthropogenic aerosols are exactly those for which there is greatest uncertainty [Myhre et al., 2013]. Hence, if we had chosen different realizations of these forcings to apply, then the resulting RF would have differed. Here we limit further specific discussion to issues around each of these three forcings and how they were applied.

6.1.1. Solar

While modifying the solar forcing, several possible alternative estimates of the observed evolution were found, based on different sources and different algorithms. Figure 15 shows the three main alternatives compared to the original forcings used in NorESM for CMIP5 including the final choice outlined in section 3. The Institut Pierre Simon Laplace solar forcing is identical to that of NorESM until 2009, when the previous cycle began to be repeated in NorESM. The solar forcing from SORCE is from the TM Instrument for 2003 onward. Prior to this, the SORCE forcings are based on previous reconstructions [Krivova et al., 2010] but include a small offset based on the TM measurements for the period from 2003 onward. The solar forcings from the Physikalisch-Meteorologisches Observatorium Davos (PMOD) are based on a combination of data from several instruments: ACRIM-I, ACRIM-II, ACRIM-III, VIRGO, TIM, and ERBE. The PMOD data are generally considered the most accurate set of solar forcings, given how it was constructed [Fröhlich and Lean, 1998; Fröhlich, 2006]. For this reason, and in order to stretch our Sensitivity tests since PMOD has the largest difference from the original NorESM runs, this estimate of solar forcing was used. However, driving bases for NorESM were created for all three of the solar forcings shown in Figure 15, and future work could explore such sensitivity.

6.1.2. Volcanic Forcing

Recent work by Ridley et al. [2014] that postdates the production of the ensemble and initial submission of this paper suggests that the volcanic forcings from Sato et al. used in the Reference forcings and Sensitivity forcings are biased low. This is based upon a comparison to other data sets based upon observations from several lidars and other instruments. Specifically, an assumption regarding the lowermost level in which stratospheric volcanic aerosols can reside leads to an underestimate of the stratospheric burden in models to high latitudes, particularly in winter. This requires a reassessment of how to apply this forcing so the stratospheric aerosol loading base pressure level varies seasonally and latitudinally. This is not the case in either of the current ancillaries in NorESM, and for that reason the other ensemble member GCMs and ESMs, which load above a fixed pressure level with no global seasonal redistribution, all have a higher forcing due to volcanic aerosol effects. This is an important point, as it shows how the sensitivity of the model is impacted by different volcanic aerosol forcings. However, given the uncertainty in volcanic forcing, it is important to consider a range of possible forcing scenarios in future studies.

6.1.3. Aerosols

First, the estimate of forcing due to the applied anthropogenic aerosols in section 3 is incomplete as it does not account for the direct effect of aerosols on longwave radiation, which is assumed to be small. Further, while it accounts for changes in the cloud liquid water path, it does not account for the changes in cloud fraction or the indirect effect of aerosols in cold or mixed-phase clouds (the latter having a low level of understanding) [e.g., see Rinkevæg et al., 2013].
The aerosol issues as they pertain to the present analysis were highlighted in many areas of the recent Fifth Assessment Report with, in particular, a whole chapter concerned with aerosols and clouds [Boucher et al., 2013]. The radiative effects of aerosols are an area of active research as highlighted by recent high profile work on the issue by, for example, Shindel [2014] and Stevens [2015]. Relevant principal aspects to the present study are as follows. Hartmann et al. [2013] concluded that while regional changes were apparent in regions with sufficient satellite records, these are grossly incompleteness and that confidence in satellite-based aerosol optical depth records is low. Myers et al. [2013] highlighted large uncertainty in historical aerosol forcings and their radiative effects, and that these continued to be the primary source of uncertainty in historical forcings. Khim et al. [2013, section 11.3.6.1] cautioned that the aerosol emission-RCP scenarios in the short term may not be realistic, even to the shortest time frame. They further note that that when and where the aerosols are emitted may be important, so a consideration of the effect of aerosols should not necessarily be considered in terms solely of the global mean RF. Postdating the report, Shindel [2014] suggests that the efficacy of aerosol forcings is likely to be invariant in space and time. Collins et al. [2013] highlights a roughly equal split between models that prescribe aerosol burdens directly or calculate them interactively based upon prescribed emissions (see their Table 2.1). The NorESM model considered herein uses prescribed emissions.

Given that NorESM is configured such as to derive aerosol concentrations interactively from prescribed driver ancillaries, the use of direct observationally based aerosol burdens was precluded from the outset. Historically, emission-based scenarios depend upon the veracity of national emissions inventories which cannot be taken as a given. In particular, national emissions inventories tend to concern industrial scale pollution rather than domestic source pollution. This may lead to significant underestimation of aerosol emissions in countries such as China and India. These are also considerable uncertainty as to how the correct emissions are and the lack of adequate observational capabilities for speciated aerosols globally combined with their short lifetime excludes the possibility of a budget closure being possible from any LIGHGs. During the experimental setup, an additional possible set of emissions drivers were considered. However, this yielded burdens of the speciated aerosols that were substantially and systematically different to the CMIP runs in all cases. Use of such ancillaries, while they may well be closer to the truth, would have led to large-scale shock terms being applied to the model in terms of the RF. We carefully considered the possibility of the aerosol drivers we used in Sensitivity. Such an assumption is limited by the questionable veracity of the RCP scenarios’ short-term aerosol behavior [Khim et al., 2013]. Figure 16 compares the reference and Sensitivity burdens to those for all four RCP scenarios. Under the assumption that the RCP scenarios describe plausible decadal scale variation, the applied Sensitivity forcings do not appear to be grossly unrealistic, although the POM concentrations are somewhat outside the RCP bounds.

We conclude that of the various forcings addressed in this work, both the emissions and the concentrations of tropospheric aerosols have the greatest uncertainties associated with them, which can arise from both the uncertainties in observations/emissions of aerosols and from the RCP scenarios due to uncertainties in the prediction of future emissions. One study has shown that increased sulfur emissions in Asia are important at least regionally, since the emitted sulfur counteracts the anthropogenically driven warming [Neely et al., 2013]; however, recent work by Murphy [2013] suggests that the shift of pollution from Europe to South East Asia has had little effect on the clear-sky radiative forcing. The uncertainties in aerosol as and sedimentary aerosols are considered in the model ensemble with the model run that includes the Intergovernmental Panel on Climate Change’s (IPCC) Climate Change Assessment Report (AR5) scenario.
ensembles are still not sufficiently large to sample the entire solution space of possible realizations for this model to fully explore the effects of imperfect knowledge of the forcings over the period (section 6.1). Given the chance to repeat the experiment, there are certain things that we would do differently.

If the OHC initial state is important for correctly sampling the solution space of the internal variability [Meehl et al., 2014], then starting from just three states of OHC might not be sufficient unless one of these states by chance was reasonably proximal to the 1980s OHC state. Furthermore, it may take longer than the 18-year spin-up used here to attain reasonable spread in OHC states in the abyssal oceans. It would be preferable to start from a broader range of initial OHC states to better assess whether the ocean initial state is potentially important.

With regards to ensemble design, when the runs were started, we had not diagnosed the RF of the two ensembles. Given the large changes in the forcings from any of the forcings, it was expected that the RF would differ substantially globally and certainly regionally, and hence, two distinct pathways would suffice to answer at least to consider what role forcing may play. In hindsight, the resulting global RF is somewhat similar (section 3.6). It may well be the case that all of the applied forcings in Sensitivity are reasonably proximal to the unknown truth and that the CMIP-5 runs through fortuitous calculation of means got close to this. But it would be incredibly naive to place any faith in such an inference, given the uncertainties in many of the forcings (section 6.1). If we were to repeat the analysis, we would run a somewhat larger number of
of ensemble means with ensemble members all populations so that we can better explore forcing uncertainty effects such that we have three or four different sensitivity ensemble means and hence 4 or 5 degrees of freedom in the applied forcings. Efforts to explore uncertainties in solar irradiance, address inadequacies in volcanic forcings that have only very recently become apparent, and uncertainties in aerosol forcing uncertainty could be prioritized.

Perhaps, most importantly, the use of a single model in this study ensures that the results do not suffer from the entire parameter space of realization of the hiatus obtainable by climate models, in general, under the range of forcings we considered. The sensitivity forcings used in this project have been made available through supplemental material both in the interest of transparency in our work and in the hope that they may be of use to other modeling groups interested in performing comparable studies of their own. We would also encourage groups to improve upon our experimental design.

7. Summary

Various forcings in the Norwegian Earth System models were modified to bring them in line with new observational or model-based estimates. These were applied to a 30-member sensitivity ensemble covering the period of 1980 until 2012. A similar 30-member reference ensemble was created using the CMIP5 historical forcings extended with RCP8.5. Exploratory analysis has revealed that the inclusion of these alternative forcings did not greatly alter the NSMO variability or Arctic sea ice extent between the reference and sensitivity runs. Ocean heat content did show an increased positive trend in response to the sensitivity forcings.

In summary, the observational data sets, and both ensemble mean behavior and individual ensemble members of the ensemble bles are discussed.

References


OUTDEN ET AL. INVESTIGATING RECENT HIATUS, 1


