Use of Short-Range Forecasts to Evaluate Fast Physics Processes Relevant for Climate Sensitivity

K. D. Williams, A. J. Hewitt, and A. Bodas-Salcedo

Abstract  The configuration of the Met Office Unified Model being submitted to CMIP6 has a high climate sensitivity. Previous studies have suggested that the impact of model changes on initial tendencies in numerical weather prediction (NWP) should be used to guide their suitability for inclusion in climate models. In this study we assess, using NWP experiments, the atmospheric model changes which lead to the increased climate sensitivity in the CMIP6 configuration, namely, the replacement of the aerosol scheme with GLOMAP-mode and the introduction of a scheme for representing the turbulent production of liquid water within mixed-phase cloud. Overall, the changes included in this latest configuration were found to improve the initial tendencies of the model state variables over the first 6 hr of the forecast, this timescale being before significant dynamical feedbacks are likely to occur. The reduced model drift through the forecast appears to be the result of increased cloud liquid water, leading to enhanced radiative cooling from cloud top and contributing to a stronger shortwave cloud radiative effect. These changes improve the 5-day forecast in traditional metrics used for numerical weather prediction. This study was conducted after the model was frozen and the climate sensitivity of the model determined; hence, it provides an independent test of the model changes contributing to the higher climate sensitivity. The results, along with the large body process-orientated evaluation conducted during the model development process, provide reassurance that these changes are improving the physical processes simulated by the model.

Plain Language Summary  Climate sensitivity is a leading order measure of the climate system. The latest Met Office model has a higher climate sensitivity than its predecessor and many other models, so warrants additional tests. Here we follow a published method to test in weather forecast mode, model changes contributing to the higher climate sensitivity. The model changes increasing the climate sensitivity are found to improve the short-range weather forecast and reduce the error growth over the first few hours of the forecast which is a measure of the error in the local physical processes. This increases our confidence that the model changes contributing to the higher sensitivity are improving the physical realism of the model.

1. Introduction

Equilibrium climate sensitivity, defined as the global mean temperature change at equilibrium resulting from a doubling of carbon dioxide concentration, is a leading order measure of the climate system. The higher the climate sensitivity, the larger the global temperature response, and other related parameters such as the hydrological cycle, to an external forcing.

Andrews et al. (2019) show that the climate sensitivity in HadGEM3-GC3.1 (Hadley Centre Global Environmental Model 3 - Global Coupled configuration 3.1 Williams et al., 2018) is higher than the previous HadGEM3-GC2 configuration (Williams et al., 2015), increased from 3.2 to 5.5 K as estimated using the method of Gregory et al. (2004). Given that HadGEM3-GC3.1 is being used for climate projections within CMIP6 (Coupled Model Intercomparison Project 6) and that its climate sensitivity is higher than the Intergovernmental Panel on Climate Change (IPCC), likely range of 1.5–4.5 K (Collins et al., 2013), there is naturally considerable scrutiny of the model and the processes leading to the higher climate sensitivity.

Bodas-Salcedo et al. (2019) identify the change in the atmospheric feedbacks contributing to the higher climate sensitivity in HadGEM3-GC3.1 compared with HadGEM3-GC2 as being almost entirely associated with two model changes: 1) The introduction of a new modal aerosol scheme (Global Model of Aerosol Processes modal scheme; GLOMAP-mode) with associated changes to aerosol activation, replacing the...
previous CLASSIC (Coupled Large-scale Aerosol Simulator for Studies In Climate) aerosol scheme and
2) the introduction of a scheme for representing the turbulent production of liquid water in mixed-phase
cloud. Andrews et al. (2019) show that a stronger sea ice feedback and larger greenhouse gas forcing also
contribute to the higher climate sensitivity. The stronger greenhouse gas forcing is due to improvements to
the treatment of gaseous absorption which result in better comparisons with line-by-line models (Walters
et al., 2019).

Prior to being accepted, atmospheric changes to the Unified Model (UM Cullen, 1993) are normally tested in
an AMIP (Atmosphere Model Intercomparison Project) simulation and set of numerical weather prediction
(NWP) case studies, with acceptable performance being required in both. These tests were performed with
the mixed-phase scheme, and the AMIP test was performed with GLOMAP-mode; however, at the time,
the code to allow NWP testing of the GLOMAP-mode aerosol scheme was not available. This has now been
rectified.

Both of the atmospheric model changes now identified as leading to a stronger positive feedback were sub-
ject to a wide-ranging assessment when they were included into the model, as was the HadGEM3-GC3.1
configuration as a whole. This included evaluation of the mean climate, variability, and process-orientated
evaluation such as regime compositing and detailed comparisons with observations using satellite simula-
tors (Bodas-Salcedo et al., 2011). Some of this assessment is presented in Walters et al. (2019) and Williams
et al. (2018).

Rodwell and Palmer (2007) propose an additional test for climate models based on short-range NWP exper-
iments. While any one NWP forecast will evolve from its analysis state, over a large number of forecasts
any systematic drift early in the forecast is likely to be due to errors in the representation of local physical
processes in the model. Hence, Rodwell and Palmer (2007) argue that evaluation of the average initial ten-
dency error of model state variables from an analysis consistent with that model provides a measure of the
error in the simulated physical processes in the model. Since the same physical parametrizations determine
the feedbacks in climate change simulations, they argue that a climate model with a large initial tendency
might be less trustworthy for its climate change projections since the large initial tendency would indicate
an error in the simulated physical processes. The initial tendency timescale evaluated (6 hr) is chosen to be
sufficiently short that significant feedback from the dynamics is unlikely to occur. This type of experiment
requires a set of NWP simulations with cycling data assimilation as might be run in an operational NWP
center (in Met Office nomenclature, this is known as an “NWP trial” and is a notably more complicated
experiment than the “NWP case studies” used in the standard tests which are initialized from an existing
independent analysis).

In this short paper we present results from a set of experiments following the methodology of Rodwell and
Palmer (2007) for the model changes identified by Bodas-Salcedo et al. (2019) as contributing to the higher
climate sensitivity in HadGEM3-GC3.1. It should be noted that these tests have been conducted after the
model was frozen and climate sensitivity obtained and published, hence provide a completely independent
assessment of the model.

Following a description of model components being tested and the experimental design in section 2, we
provide results from the experiments in section 3. Discussion and conclusions are in section 4.

2. Experimental Design

The configuration of the UM used in this study is Global Atmosphere version 7 (GA7) which is the atmo-
sphere (and land) component of the coupled configuration HadGEM3-GC3.1 being used for CMIP6. GA7 is
fully described by Walters et al. (2019). In these experiments we revert two components to the earlier GA6
(Walters et al., 2017) configuration. Bodas-Salcedo et al. (2019) show that the same components can be iden-
tified as being responsible for the higher climate sensitivity in GA7 regardless of whether they are added to
GA6 or removed from GA7.

GA7 uses the GLOMAP-mode aerosol scheme (Mann et al., 2010), hereafter referred to as simply GLOMAP.
The implementation of GLOMAP in GA7 is fully described by Mulcahy et al. (2019). The scheme simulates
speciated aerosol mass and number in four soluble modes covering the submicron to supermicron aerosol
size ranges (nucleation, Aitken, accumulation, and coarse modes) as well as an insoluble Aitken mode.
The aerosol species represented in GLOMAP in this configuration are sulfate, black carbon, organic carbon,
and seasalt. It replaces the CLASSIC aerosol scheme (Bellouin et al., 2011) used in GA6. In climate simulations of the physical model, the aerosol concentrations are calculated interactively from given emissions. However, the initial aerosol concentrations for a given forecast are typically unknown, and the computational cost of fully prognostic aerosols in NWP is prohibitively expensive; hence, for the NWP simulations presented here we use a seasonally varying climatological aerosol concentration from the respective AMIP simulation (i.e., concentrations for NWP GLOMAP experiments are from an AMIP simulation using interactive GLOMAP, and concentrations for NWP CLASSIC experiments are from an AMIP simulation using interactive CLASSIC). This traceable approach of using climate model climatologies in the NWP simulation is currently used in Met Office operational weather forecasts. Mineral dust is simulated using CLASSIC in both GA6 and GA7. The direct and indirect effects of aerosols are calculated fully interactively in the NWP experiments. With GLOMAP aerosols, the cloud droplet number concentrations are calculated using the activation scheme of West et al. (2014), which is based on the scheme of Abdul-Razzak and Ghan (2000). With CLASSIC, “activation” is an empirically based parameterization of droplet number as a function of aerosol number following Jones et al. (2001).

A representation of the turbulent production of liquid water in mixed-phase cloud was introduced in GA7. The scheme is based on Field et al. (2014) and its introduction in the UM is described by Furtado et al. (2016) and Furtado (2018). The scheme analytically solves the dynamics of supersaturation fluctuations in turbulent air motions, exchange of air between the cloud and its environment, and the depletion of supersaturation by microphysical growth of the ice phase. This is used to calculate a probability distribution of supersaturation, with the cloud water content and fraction being calculated as moments of this distribution.

Four NWP trials have been conducted:
1. GA7 (which includes GLOMAP aerosol concentrations, with West et al. (2014) activation and turbulent production of mixed-phase cloud).
2. GA7 but without the turbulent production of mixed-phase cloud liquid water.
3. GA7 but using the CLASSIC aerosol concentrations (and Jones et al., 2001 activation).
4. GA7 but using the CLASSIC aerosol concentrations and without the turbulent production of mixed-phase cloud liquid water.

A three-month trial comprising four forecasts per day with fully cycling data assimilation (DA) has been run for each experiment at a horizontal resolution of N320 (40 km in the midlatitudes), although our experience is that results will be qualitatively similar at different resolutions. The Met Office hybrid-4Dvar (four-dimensional variational) DA system is used, so the observations are introduced over a time window of up to 3 hr either side of analysis time. As well as providing a more accurate analysis, this approach minimizes any model adjustment early in the forecast. The DA is the same as used operationally by the Met Office except that in the trials performed here, the DA is uncoupled to the ensemble. The trial covers the period 1 December 2017–28 February 2018 inclusive, a DJF trial being chosen since Bodas-Salcedo et al. (2019) show that the increased climate sensitivity in GA7 is mainly through a shortwave feedback in the Southern hemisphere and the austral summer is when insolation is highest.

This is the same experimental design as Rodwell and Palmer (2007), with the only difference being that they make changes to their linear adjoint within their 4D-var DA system consistent with the forward model for each experiment. The Met Office DA system uses a perturbation forecast (PF) model with highly simplified physics rather than a true linear adjoint. Operationally the PF model is only updated occasionally rather than with every forward model change, and since there is no explicit representation of aerosols or mixed-phase clouds in the PF model, no changes are made to the PF model in the experiments presented here.

3. NWP Evaluation of Fast Physics Processes

We first assess the experiments using standard NWP dynamical performance measures. These are some of the main scores exchanged between centers under the WMO CBS (World Meteorological Organization Commission for Basic Systems) protocol (e.g., https://apps.ecmwf.int/wmolcdnv/). We use the GA7 configuration (including GLOMAP and the mixed-phase scheme) as the control and “scorecards” showing the impact of reverting components to GA6 are presented in Figure 1. It can be seen that either reverting the aerosol scheme to CLASSIC or removing the turbulent production of mixed-phase cloud liquid water has a detrimental impact on the forecast. This is particularly the case for the forecast root mean square error of
Figure 1. NWP “scorecard” showing relative difference in RMSE against SYNOP observations compared with GA7 when (a) reverting from GLOMAP to CLASSIC aerosol, (b) removing the mixed-phase scheme, and (c) both reverting to CLASSIC and removing the mixed-phase scheme. Increased RMSE are red. The size of the colored box represents the fractional change with a 2% change filling the outline gray box. Variables are pressure at mean sea level (PMSL), 500-hPa geopotential height (H500), winds at 850 and 250 hPa (W850, W250) for the northern hemisphere (NH, 20–90°N), tropics (Trop, 20°S–20°N), and Southern hemisphere (SH, 20–90°S). These are shown against forecast lead time (hours).
Figure 2. (a) Vertical profile of average temperature tendency for the first 6 hr of all the NWP forecasts for each experiment. (b) Vertical profile of average cloud liquid water content 6 hr into each forecast. Both are for the region 25–70°S. Model level data (model levels up to 10 km) are plotted.

500-hPa geopotential height (H500 RMSE) over the Southern hemisphere (here 20–90°S) across all forecast lead times out to 5 days (the length of forecasts run in these experiments). The experiment both reverting the aerosols to CLASSIC and removing the mixed-phase scheme shows a larger degradation to the forecast implying that the impact of the two changes are, to some extent, independent. The deterioration in the forecast is significant and up to a 2% increase in H500 RMSE in the experiment removing both GLOMAP and the mixed-phase scheme. This is notable, although in our experience other physics changes can have a larger impact than this. Verification of each experiment in Figure 1 is against its own analysis, however qualitatively the same results are obtained when verified against SYNOP observations (not shown as these are relatively sparse in the Southern hemisphere).

If a model were a perfect representation of the real world, and the initial analysis perfectly represented the current state, then the 6-hr forecast would be an accurate representation of the state of the atmosphere at that 6-hr point, and the DA would need to make no change in order to produce the next analysis (in DA nomenclature, the analysis increment to the model background would be zero). In reality, the model, observations, and analyses all contain errors; however, we assume here that the analysis is the best representation of the current state of the atmosphere and the errors are smaller than the model forecast errors (even over the Southern Ocean, there is a wealth of satellite data assimilated in modern DA systems). Systematic errors in the model will tend to evolve the forecast locally in a particular direction (e.g., always warming in a particular meteorological situation), so over a large number of forecasts the difference between the forecast used
Figure 3. DJF zonal mean difference in cloud liquid water content between GA6 and GA7 for a 27-year AMIP simulation. GA7 minus GA6 is plotted.

The 6-hr temperature tendency, as an average across all the forecasts in the trial for this region, is for a cooling in the upper troposphere and warming in the lower troposphere in all the experiments (Figure 2a). The impact of the model changes considered in this paper is confined to a layer below 3 km. Just above the cloud layer at around 1.5 km, the GLOMAP aerosol scheme and the mixed-phase cloud scheme reduce the warming tendency which is beneficial. These components do not completely eliminate the temperature error as there remains a warm tendency at around 3 km and also near the surface which is actually made slightly worse when CLASSIC aerosols are replaced by GLOMAP. Nevertheless, GLOMAP and the mixed-phase scheme may be regarded as improving the temperature tendency profile at the top of the boundary layer and above which results in a reduced bias in 500 hPa geopotential height and is likely the cause of the improved forecast RMSE with these components noted in Figure 1.

The cooling just above the boundary layer can be understood as increased radiative cooling from the larger amount of cloud liquid water content (Figure 2b). This is entirely consistent with the higher cloud liquid water seen in the boundary layer when comparing GA6 and GA7 AMIP experiments (Figure 3). Bodas-Salcedo et al. (2016) find that there is an abundance of super-cooled liquid cloud over the Southern Ocean suggested by satellite observations and models generally underestimate this, with GA6 (and to a lesser extent GA7) being no exception. The mixed-phase scheme was developed specifically to target this issue with the cloud liquid water being calculated theoretically for steady state given the grid box conditions, including the turbulent kinetic energy, and this amount of cloud liquid water effectively imposed. Therefore the increased cloud liquid water would be fully expected from this change. As discussed by Bodas-Salcedo et al. (2019), switching to GLOMAP can affect the number of particles activated into cloud drops through several mechanisms. Overall the introduction of GLOMAP increases the aerosol concentration in the Southern hemisphere leading to smaller cloud drops and reduced effective radius. It would be expected that this would reduce auto-conversion, increasing the cloud liquid water as found in Figure 2b. However, Bodas-Salcedo et al. (2019) only find small differences in liquid water path in an AMIP simulation when switching to GLOMAP. It is possible that the increased liquid water seen early in the initialized simulations is the rapid effect expected, but on longer timescales, other processes feed back to counter the increased liquid water path. Cloud ice water content is also slightly increased in these simulations when the mixed-phase scheme is introduced, but by a much smaller amount (not shown). There is little change in the overall specific humidity tendency.

Rodwell and Palmer (2007) propose metrics to compare models using the initial tendency approach as being the mass-weighted vertically integrated tendency for each of four state variables: temperature, specific humidity, the zonal, and meridional components of wind. For the present set of trials, it can be seen that the mixed-phase scheme is a clear improvement (Table 1). The introduction of GLOMAP gives little change in the T tendency since the detrimental warming below cloud offsets the beneficial cooling above. However, the wind metrics are clearly improved suggesting these might be more sensitive to temperature errors in the upper boundary layer and free troposphere. Inclusion of the turbulent production of water in mixed-phase cloud is an improvement on all the metrics. The differences between the metrics for the experiments in Table 1 are smaller than the differences between the perturbed parameter experiments undertaken by Rodwell and Palmer (2007), reflecting that the GA7 changes are modest improvements. However, given that the increased tendency metric in the experiment removing both changes from GA7 is, in every case, at least as
Table 1

|                | T (K day\(^{-1}\)) | q (g kg\(^{-1}\) day\(^{-1}\)) | u (m s\(^{-1}\) day\(^{-1}\)) | v (m s\(^{-1}\) day\(^{-1}\)) | Overall score |
|----------------|--------------------|---------------------------------|-------------------------------|-------------------------------|---------------|
| GA7            | 0.0820             | 0.0160                          | 0.188                         | 0.185                         | 0.79          |
| GA7 minus GLOMAP | 0.0821             | 0.0160                          | 0.212                         | 0.193                         | 0.99          |
| GA7 minus mixed phase | 0.0897         | 0.0163                          | 0.204                         | 0.187                         | 1.01          |
| GA7 minus both | 0.0897             | 0.0163                          | 0.217                         | 0.193                         | 1.02          |

Note. The right column provides an overall score which is the normalized average of the four parameters (normalized by the average score for each variable). In all columns, lower values imply smaller errors. Unlike Rodwell and Palmer (2007), who use 11 representative levels, we include all model levels from the surface to level 50 (around 75 hPa/18 km).

large as the experiment reverting one change and in the same direction on each model level, we suggest that the changes in the metrics for the experiment removing both are significant. Rodwell and Palmer (2007) also propose an overall score which is the average of the scores for the four variables, each normalized by the average score for that variable across the models being considered. This overall score is in the right-hand column with a lower number implying smaller overall errors. GA7 achieves the best (lowest) score of the four experiments and while averaging the four variables may be considered to be placing a lot of weight on the wind components, we note that GA7 has the best, or joint best, score in each of the variables.

Shortwave cloud radiative effect (SCRE), defined as the difference between the net downward all-sky shortwave at the top of the atmosphere and that in cloud-free conditions, is a leading measure of the cloud shortwave radiative properties. Bodas-Salcedo et al. (2019) show that it is through changes in the SCRE that GA7 has the higher climate sensitivity. Given the known importance for the climate change response, 20-year mean geographical maps of SCRE from AMIP simulations are among the key diagnostics assessed during the model development process. Here, we compare the day 1 mean forecast SCRE for the DJF 2017/18 NWP trials with the same season from observations (Figure 4). While not perfect, the GA7 forecasts are clearly the closest fit to the CERES observations with a stronger (more negative) SCRE over the Southern Ocean, consistent with the higher cloud water content making the cloud more reflective. Bodas-Salcedo et al. (2019) show that GLOMAP also contributes to the enhanced SCRE in climate simulations through changes in the droplet effective radius. It is likely that the same is true in these experiments although diagnostics of effective radius were unavailable in the NWP trials. Hence, the improved SCRE in Figure 4 is probably a combination of the increased cloud liquid water content discussed above and smaller effective radius.

Figure 4. Day 1 average zonal mean SCRE for the NWP experiments and the corresponding DJF 2017/18 satellite observations from Clouds and the Earth’s Radiant Energy System (CERES Loeb et al., 2005) for the Southern hemisphere.
4. Conclusions and Discussion

The NWP assessment presented in this paper has provided an additional test of the atmospheric model changes in GA7 known to contribute to the increased climate sensitivity compared with the previous GA6 configuration. Other factors contributing to the higher sensitivity are the increased greenhouse gas forcing which is due to a radiation improvement leading to better comparison with line-by-line models and a slightly stronger sea ice feedback (the 6-hr initial tendency method presented here not being an relevant test for the slower sea ice processes). The atmospheric processes investigated in this paper contribute to the larger climate sensitivity through changes in shortwave cloud radiative feedbacks. We find that the stronger present-day SCRE in climate simulations can also be seen in the NWP trials and is due to increased cloud liquid water when the GLOMAP aerosol scheme replaces the former CLASSIC scheme and when a representation of the turbulent production of mixed-phase cloud liquid water is included. The increased radiative cooling from these clouds reduces a warming tendency just above cloud top which is present throughout the lower troposphere as the model evolves over the first 6 hr of the forecast. These in turn lead to improved predictability over the 5-day NWP forecasts, for example, in the H500 RMSE. Despite the improvements in GA7, there remains a tendency for lower tropospheric warming and upper tropospheric cooling over the Southern hemisphere and future model changes should aim to reduce this.

We calculate overall metric scores for the four NWP experiments. GA7, including both GLOMAP and the mixed-phase scheme, achieves the best score. Rodwell and Palmer (2007) suggest that this metric could be used to assign a weight to climate models. We consider it a valuable addition to the wider basket of assessment measures and process-orientated diagnostics used to evaluate climate models. As these NWP trials were only performed after the model was frozen and climate sensitivity obtained, in this case the metrics do provide an important independent assessment of the model changes.

The changes leading to the higher climate sensitivity in GA7 compared with GA6 are found to be beneficial in the NWP tests. Together with the process-orientated assessments conducted during the model development cycle and physical basis of the changes, the results provide reassurance that the changes responsible are improving the realism of the model. More generally, we can find no evidence that the increase in climate sensitivity is erroneous. It is, of course, still possible that the climate sensitivity of GA7 is in error due to other existing processes in the model being inadequately represented, and thus resulting in incorrect feedbacks. However, all of the process-related work undertaken to date suggests that GA7 has a good representation of the key climate feedbacks effecting climate sensitivity.

The initial tendency method is well suited to climate modeling systems which also have a data assimilation component. An alternative approach for testing climate models in weather forecasting mode is transpose-AMIP (Williams et al., 2013), where the model is initialized from an existing, high quality, NWP analysis. However, the adjustment when initializing from an inconsistent analysis means that the first few hours of a forecast may need to be disregarded (e.g., Ma et al., 2014); hence, the initial tendency methodology typically can not be followed using the transpose-AMIP approach. That said, experience at the Met Office is that the performance of day 2–5 forecasts when initialized from an independent analysis can often be a reasonable (although not perfect) guide to the performance in a trial with cycling data assimilation, especially in cases where the impact is larger (e.g., Walters et al., 2017). We therefore suggest that transpose-AMIP style experiments would provide a useful additional model test for climate models if a consistent data assimilation system was unavailable, and hence, the initial tendency method could not be followed.
Bodas-Salcedo, A., Mulcahy, J. P., Andrews, T., Williams, K. D., Ringer, M. A., Field, P. R., & Elsaesser, G. S. (2019). Strong dependence of atmospheric feedbacks on mixed-phase microphysics and aerosol-cloud interactions. *Journal of Advances in Modeling Earth Systems, 11*, 1735–1758. https://doi.org/10.1029/2019MS001688

Bodas-Salcedo, A., Webb, M. J., Bony, S., Chepfer, H., Dufresne, J. L., Klein, S., et al. (2011). COSP: Satellite simulation software for model assessment. *Bulletin of the American Meteorological Society, 92*(8), 1023–1043. https://doi.org/10.1175/2011BAMS2856.1

Collins, M., Knutti, R., Arblaster, J. M., Dufresne, J., Fichefet, T., Friedlingstein, P., et al. (2013). In T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, et al. (Eds.), *Long-term climate change: Projections, commitments and irreversibility, in Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 1029–1136). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781107415324.024

Cullen, M. J. P. (1993). The unified forecast/circulation model. *Meteorological Magazine, 122*, 81–94.

Field, P. R., Hill, A. A., Furtado, K., & Korolev, A. (2014). Mixed-phase clouds in a turbulent environment. Part 2: Analytic treatment. *Quarterly Journal of the Royal Meteorological Society, 149*, 870–880. https://doi.org/10.1002/qj.2175

Furtado, K. (2018). Mixed-phase clouds, chap. Chapter 8 - Subgrid representation of mixed-phase clouds in a general circulation model (pp. 185–214). Constantin Andronache: Elsevier. https://doi.org/10.1016/B978-0-12-810549-8.00008-8

Furtado, K., Field, P. R., Boutle, I. R., Morcrette, C. J., & Wilkinson, J. M. (2016). A physically-based, subgrid parametrization for the production and maintenance of mixed-phase clouds in a general circulation model. *Journal of the Atmospheric Sciences, 73*, 279–291. https://doi.org/10.1175/JAS-D-15-0021.1

Gregory, J. M., Ingram, W. J., Palmer, M. A., Jones, G. S., Stott, P. A., Thorpe, R. B., et al. (2004). A new method for diagnosing radiative forcing and climate sensitivity. *Geophysical Research Letters, 31*, L03205. https://doi.org/10.1029/2003GL018747

Jones, A., Roberts, D. L., Woodage, M. J., & Johnson, C. E. (2001). Indirect sulphate aerosol forcing in a climate model with an interactive sulphur cycle. *Journal of Geophysical Research, 106*, 20,293–20,310.

Klinker, E., & Sardeshmukh, P. D. (1992). The diagnosis of mechanical dissipation in the atmosphere from large-scale balance requirements. *Journal of the Atmospheric Sciences, 49*, 608–627. 

Loeb, N. G., Kato, S., Loukachine, K., & Manalo-Smith, N. (2005). Angular distribution models for top-of-atmosphere radiative flux estimation from the Clouds and the Earth’s Radiant Energy System instrument on the Terra satellite. Part I: Methodology. *Journal of Atmospheric and Oceanic Technology, 22*, 338–351. https://doi.org/10.1175/JTECH1712.1

Ma, H.-Y., Xie, S., Klein, S. A., Williams, K. D., Boyle, J. S., Bony, S., et al. (2014). On the correspondence between mean forecast errors and climate errors in CMIP5 models. *Journal Climate, 17*, 1781–1798. https://doi.org/10.1175/JCLI-D-13-00474.1

Mann, G. W., Carslaw, K. S., Spracklen, D. V., Ridley, D. A., Manketelow, P. T., Chipperfield, M. P., et al. (2010). Description and evaluation of GLOMAP-mode: A modal global aerosol microphysics model for the UKCA composition-climate model. *Geoscientific Model Development, 3*, 519–551. https://doi.org/10.5194/gmd-3-519-2010

Mulcahy, J. P., Johnson, C., Jones, C., Povey, A., Sellar, A., Scott, C. E., et al. (2019). Description and evaluation of aerosol in UKESM1 and HadGEM3-GC3.1 CMIP6 historical simulations. *Geosci. Model Dev.*, 12, 1509–1524. https://doi.org/10.5194/gmd-12-1509-2019

Williams, K. D., Harris, C. M., Bodas-Salcedo, A., Camp, J., Comer, R. E., Copsey, D., et al. (2015). The Met Office global coupled model 2.0 (GC2) configuration. *Geoscientific Model Development, 8*, 1509–1524. https://doi.org/10.5194/gmd-8-1509-2015