Abstract As the stratosphere is largely characterized by its ozone abundance, the quality of the ozone field is important for a realistic representation of the stratosphere in weather and climate models. While the stratosphere is directly affected by radiative heating from ozone photodissociation, ozone abundance might also impact the representation of the troposphere since the stratosphere and troposphere are dynamically linked. In this paper, we examine the potential benefits of using ozone data from the Earth Observing System (EOS) Microwave Limb Sounder (MLS) for medium-extended range tropospheric forecasts in a current numerical weather prediction system. The global component of the Met Office Global and Regional Ensemble Prediction System is used, which is run at a resolution of N216 L85 with 24 ensemble members. We compare two scenarios of 31 day forecasts covering the same period, one with the current operational ozone climatology and the other with a monthly mean zonally averaged ozone field computed from the MLS data set. In the extreme case of the Arctic “ozone hole” of March 2011, our results show a general reduction in stratospheric forecast errors in the tropics and Southern Hemisphere as a result of the improved representation of ozone. However, even in such a scenario, where the MLS ozone field is much superior to that of the control, we find that tropospheric forecast errors in the medium-extended range are dominated by the spread of ensemble members and no significant reduction in the root-mean-square forecast errors.

1. Introduction

In recent years, there has been an increasing interest in the area of stratosphere-troposphere coupling. It is now widely acknowledged that changes in the stratosphere can have a significant impact on the troposphere through dynamical processes [e.g., Haigh et al., 2005; Polvani and Kushner, 2002; Williams, 2006; Simpson et al., 2009]. In the context of numerical weather prediction (NWP), these findings raise questions about whether or not a better represented stratosphere would improve the accuracy of weather forecasts especially in the lower troposphere [e.g., Jung et al., 2010].

Ozone (O₃), as a strong absorber of ultraviolet (UV) radiation, provides diabatic heating in the stratosphere and maintains the static stability in that region. Hence, the quality of ozone data used in NWP models is important for a realistic representation of the stratosphere and potentially for the accuracy of forecasts. Assimilation of ozone data into the current NWP systems should provide more accurate radiative heating rates and hence a more accurate atmospheric temperature field [De Grandpré et al., 2009]. In addition, the assimilation of satellite radiances from temperature-sounding channels that are sensitive to ozone concentration may possibly benefit from ozone assimilation. Furthermore, zonal wind analyses in the upper troposphere-lower stratosphere might also be improved by utilizing the correlation between ozone and wind [Riishøjgaard, 1996]. It has also been reported that ozone assimilation reduces lower stratospheric wind bias in NWP models [Semane et al., 2009].

Despite the potential benefits from ozone assimilation, the current Met Office operational NWP system utilizes an ozone climatology (5 years of monthly mean zonal mean ozone field on a 2.5° × 2.5° grid from Li and Shine [1995]). Mathison et al. [2007] showed that assimilating ozone reduces short-range (days 1–5) weather forecast errors in the Met Office NWP system, but the benefits were considered small compared to the extra computational cost and so ozone assimilation was not adopted operationally by the Met Office. Ozone assimilation is, however, used operationally at other centers such as the European Centre for Medium-Range Weather Forecasts [e.g., Dethof and Hólm, 2004] and the National Aeronautics and Space Administration [e.g., Stajner et al., 2008].
While it is widely accepted that ozone assimilation improves radiative heating, it remains controversial whether or not a less costly approach (e.g., using recent instead of long-term ozone climatology) is sufficient to have any positive impact on current NWP systems. For instance, while Morcrette [2003] suggests no clear benefit to temperature forecast error by using ozone analyses, Mathison et al. [2007] report an improvement in the NWP index (with significant contribution from the tropics and Southern Hemisphere [SH]) by using an alternative ozone climatology [Randel and Wu, 1999] in place of the Met Office's operational Li and Shine climatology. They also showed impacts on longer-term (up to 60 days) deterministic forecasts.

The aim of this paper is to revisit the potential benefits of improving the representation of ozone by using an alternative climatology in the Met Office NWP system. Unlike Mathison et al. [2007], we produce ensemble forecasts rather than deterministic forecasts, with a focus on forecasts of up to 31 days. The approach here is to replace the Li and Shine ozone climatology with a different ozone data set and determine whether or not it has a positive impact on temperature, wind, and surface pressure forecasts in the current Met Office NWP system. Ozone data from the Microwave Limb Sounder (MLS) were chosen for this task, and 31 day forecasts were produced for a set of chosen periods.

The structure of this paper is as follows. Section 2 describes the method used to incorporate the MLS O3 data into the Met Office NWP system and methods used to analyze the results. The results are presented in section 3. Finally, the discussion and conclusions are given in section 4.

2. Experimental Setup
2.1. The Met Office Ensemble System
The Met Office Global and Regional Ensemble Prediction System (MOGREPS) is used in this study. MOGREPS has two components: a global ensemble which produces forecasts for the whole of the globe and a regional ensemble which only covers the North Atlantic and Europe [Bowler et al., 2008]. Both the global and regional forecasts utilize the Met Office United Model (UM) and consist of 24 ensemble members. The global ensemble component is used in the paper since changes in ozone is likely to result in adjustments in large-scale dynamics. The ensemble forecast model is set up to run at a horizontal resolution of N216 (0.83° longitude by 0.55° latitude) and on 85 vertical levels which extend from the surface to 85 km in altitude.

The full dynamical response in the troposphere induced by ozone perturbations does not take place instantly. Simpson et al. [2009] suggest that the mechanism whereby tropospheric winds accelerate as a result of stratospheric heating can be separated into two stages: a slow initial acceleration which takes place in the first 20 days followed by more rapid changes. It is therefore expected that the full effect of replacing the ozone climatology in the Met Office NWP system would not be observed until at least 20 days after the perturbation is applied. In this paper we thus produce ensemble forecasts for 31 days, by extending the operational Met Office 15 day forecast system (MOGREPS-15), with the aim of better examining ozone impacts on the troposphere in this 20–31 day period.

2.2. Selected Experimental Periods
Some regional impact of changes in ozone on the temperature and dynamics is expected over much of the year. In the Southern Hemisphere, links between Antarctic ozone depletion and the troposphere have been shown in winter and spring (e.g., Ndarana et al. [2012] and Orr et al. [2012], respectively). In the Northern Hemisphere winter, changes in the polar vortex are associated with changes in surface pressure over the North Atlantic [e.g., Baldwin and Dunkerton, 2001]. Since changes in stratospheric ozone can affect the stratospheric jets, it is plausible that they might similarly impact the surface pressure pattern. Notable perturbations in stratospheric ozone depletion may occur over the Northern Hemisphere in winter to early spring, depending on stratospheric dynamics and temperatures; a recent extreme example was the strong Arctic ozone depletion in March 2011. Based on these observations, 31 day ensemble forecasts were produced for four seasonal periods covering March 2011 to January 2012. These consisted of two 1 month periods in the Southern Hemisphere winter and spring and two in the Northern Hemisphere winter covering January 2012 and March 2011, respectively. It was found that the impact on tropospheric weather forecasts of changing the monthly ozone field was small and similar in all of these periods so, given that the difference between the two ozone fields in March 2011 was largest and more consistent over time, we choose to present results only for March 2011 in this paper.

During winter, the equator-to-pole temperature gradient allows a polar vortex to form, trapping cold air within. When the temperature is lower than 196 K, water vapor and nitric acid condense to form polar
stratospheric clouds, which allow potential ozone-depleting chemicals to be converted into their reactive form [Manney et al., 2011]. In SH, such conditions usually persist over 4–5 months and as a result, an ozone hole is observed every year over Antarctica during southern spring. In contrast, the polar winter temperature in the Northern Hemisphere (NH) is generally higher and does not facilitate ozone destroying processes so well, such that generally no ozone hole is observed over the Arctic and the NH polar vortex only persists for 2–3 months. During winter 2010–2011, however, an unusually cold and strong polar vortex, which persisted through the end of March, was observed over the Arctic. This resulted in unprecedented ozone loss over the Arctic, and ozone abundance hit a minimum in mid-March.

The current ozone climatology used operationally by the Met Office is that constructed by Li and Shine [1995]: a monthly mean zonal mean field averaged over 5 years of records (1985–1989). As the Arctic “ozone hole” in 2011 was not characterized in the Li and Shine data set, this event provides a good opportunity to study the influence of the representation of ozone in the Met Office NWP system. Figure 1 (middle) shows differences in monthly mean zonal mean ozone abundance between the MLS and Li and Shine (Figure 1 (left)) data set in March 2011. The condition persisted throughout March such that Figure 1 is also representative of daily zonal mean differences between the two data sets. The maximum ozone difference between the two is found to be 1.75 ppmv in magnitude for March 2011, located at high latitudes in the NH.

The starting conditions and perturbations for the 31 day forecast ensemble members were taken from the operational Met Office analyses. For a given start date, the model was run to produce a pair of 31 day forecasts using Li and Shine (control, run LiShine) and MLS (experiment, run MLS) ozone, respectively. Since the initial conditions were calculated using the Li and Shine data set operationally, they are not in radiative
balance with the MLS run initially and temperature adjustments are expected, which may persist for around 20 days (as implied by Simpson et al. [2009], for example).

The operational setting of MOGREPS-15 allows the ozone field to be updated at the beginning of the run and periodically as the model progresses by temporally interpolating the prescribed monthly mean ozone field to the center point of the period in question. The updating of ozone in the MLS runs is done in the same manner using monthly means from the MLS data set except at the beginning of the run. The starting ozone conditions of all MLS runs are specified with their corresponding daily means computed directly from the MLS data set. This is to allow temperature adjustments at the beginning of the MLS runs.

Both the LiShine and MLS runs are compared and verified against the Met Office’s operational analysis archive. While the Li and Shine [1995] climatology was used with the UM when generating the analyses, the choice of ozone climatology plays a comparatively minor role in affecting the analysis temperatures which largely reflect the impact of the temperature observations.

For consistency checks, we have repeated the 31 day forecasts on various start dates. For the March 2011 period, the chosen start dates were 1 to 6 March 2011, inclusive. With 24 ensemble members in each run, this equates to 24 × 6 ensemble members.

A methodology similar to Mathison et al. [2007] was chosen in setting up consecutive ensemble experiments and combining them into a single data set instead of producing a large ensemble of forecasts from a single start date. The chosen method is easier to implement than to increase the ensemble size directly. Having found that the impact is low and similar for all the periods considered, we decided to conclude after the sixth start date of the March 2011 set due to constraints on time and data storage. Since we are assessing the impact of representation ozone in a current NWP system, it is desirable to keep the NWP model in its operational setting.

Note that both the Li and Shine and MLS ozone used in this study are not full three-dimensional fields and might not give an accurate ozone gradient and distribution. Nevertheless, it is evident that the March 2011 Arctic ozone hole is not represented well by the Li and Shine ozone and MLS can be considered as the more “realistic” option in this study.

### 2.3. Incorporation of MLS O₃ Data Into the Met Office NWP System

The Microwave Limb Sounder is one of the instruments on the NASA Aura satellite launched on 15 July 2004. The satellite makes around 13 Sun-synchronous orbits per day at an altitude of 705 km and provides 240 scans per orbit between 82°S and 82°N. This translates to a horizontal resolution of 1.5° or 165 km along the suborbital track. The useful altitude range of MLS ozone spans from 261 to 0.02 hPa; the vertical resolution of the data varies from 3 km (261–0.2 hPa) to 4–5.5 km (0.1–0.02 hPa). The precision is 20–50 parts per billion by volume (ppbv) in the upper troposphere-lower stratosphere and 0.1–0.02 parts per million by volume (ppmv) above 22 hPa [Froidevaux et al., 2006].

For this study, the MLS data were incorporated into the Met Office NWP system by interpolating monthly zonally averaged ozone fields on to the model grid. As there are no data poleward of 82° in the MLS data set, the ozone concentration near the pole was set to the value observed at 82°.

The UM requires atmospheric ancillary data like ozone to be stored on a hybrid height coordinate instead of pressure levels. The altitude to which each of the pressure levels corresponds is approximated by utilizing the hydrostatic equation and the air temperature at 12 Z of the day under consideration. Using the transformed coordinates, the ozone data are interpolated onto the desired model levels. However, the useful range of MLS ozone spans only from 261 to 0.01 hPa. Values for missing levels were supplied by the Li and Shine ozone climatology. Under these considerations, only data from the following pressure levels (in hPa) are taken from the MLS ozone data set: 217, 179, 147, 121.5, 100.5, 83.00, 68, 56.5, 46.5, 38.5, 31.5, 25.5, 20.5, 15, 10, 6.85, 4.8, 3.25, 2.2, 1.5, 1, 0.7, 0.5, 0.33, 0.22, 0.15, and 0.06. No blending between the two ozone data sets is performed as the ozone concentration in the troposphere is negligible compared to that in the stratosphere. The MLS-based ozone distribution used for the March 2011 case study is shown in Figure 1 relative to the Li and Shine climatology.

Note that only ozone data within 3 h of 12 Z are used. This results in a slight difference in the monthly mean zonal mean ozone profile compared to the case in which all available data are considered. Figure 2 (right) shows that this results in 5–10% difference in the zonal mean values of ozone around the tropopause.
where the ozone abundance is low. Using partial data also makes a 5% difference near the poles in the middle-upper stratosphere, where ozone abundance is assumed to be uniform poleward of 82°. Differences between the two elsewhere are small, particularly in regions where ozone is abundant. Even though the difference is small, the MLS ozone profile used in this paper should only be considered as a quasi-zonal mean.

3. Results for March 2011

As mentioned in section 2.2, March 2011 is of particular interest because of the observed "ozone hole" over the Arctic. For all the time-averaged plots, we choose to retain only data between 21 and 31 March 2011 in each of the six ensemble forecasts as described in section 2.1. Although this implies that we are averaging a slightly different forecast range over each ensemble forecast (e.g., \(T + 480\) to \(T + 720\) for the forecast starting 1 March, \(T + 456\) to \(T + 696\) for 2 March, and so on), the overall result is very similar to that obtained by averaging over the same forecast range using a forecast ensemble (not shown) due to the extent of ensemble spread beyond \(T + 240\). The operational analyses from the Met Office will serve as the reference from which the quality of the forecasts and the relative significance in impact of the changes in ozone climatology will be assessed. Details of the calculation of statistical significance in the difference between the cases appear in Appendix A.

Figure 3 shows the temperature differences for MLS-LiShine. Consistent with Figure 1, the temperatures in the lower stratosphere for the MLS case are generally lower than those in LiShine (Figure 3e) due to diminished absorption of UV. The statistically significant temperature differences in this region are fairly

---

**Figure 2.** (left) Zonal mean MLS ozone for March 2011 using only data within 12 ± 3 Z (ppmv) and (right) the percentage difference with an equivalent MLS ozone profile computed using all available data.
uniform in the meridional direction, ranging from \(-0.5\) K to \(-1.5\) K. The signal does not extend down to the troposphere except between 50°N and 70°N.

Compared with the analysis, the MLS case overestimates air temperature in the lower stratosphere by 1–2 K in both the tropics and subtropics (Figure 3c). It can be observed from Figure 3a that LiShine further
overestimates the temperature in that region. A possible source for the overestimations with LiShine and MLS could be model errors. Poleward of NH subtropics, the temperature differences (MLS-analysis, Figure 3c) range from 0 K at 40°N to ~3 K at the North Pole in the lower stratosphere. The tropospheric temperatures in MLS are also generally lower than those of the analysis with differences being statistically significant at all latitudes.

In summary, the differences in zonal mean temperature in the cases employing the MLS and LiShine ozone fields (Figure 3e) are generally small compared to the forecast errors (Figures 3a and 3c); not a surprising result for an ozone-related study at a forecast range of 30 days. It is important to note, however, even though of small magnitude, that the difference in temperature between MLS and LiShine is statistically significant, especially in the stratosphere. In the following, we will continue to explore whether this holds true for tropospheric winds and sea level pressure (SLP) before discussing the two cases' respective forecast errors in more details.

The response of zonal mean zonal wind is shown in the right-hand side of Figure 3. The thermal wind relation requires any altered meridional temperature gradient to be balanced by a corresponding change in zonal wind. Associated with the errors in the meridional temperature gradient (Figure 3c) the MLS-analysis wind field exhibits a strong banded structure as shown in Figure 3d. In the case of MLS-LiShine, the temperature anomaly pattern in Figure 3e is weaker and more uniform meridionally. As a result, even though statistically significant in midlatitudes, the magnitude in zonal wind differences between MLS and LiShine (Figure 3f) is about 10 times smaller than that of their forecast errors (Figures 3b and 3d).

Figure 4a shows SLP differences for MLS-LiShine. In the NH, only a small portion is found to be statistically significant over the Norwegian Sea and the area to the northwest of Japan. In the SH, a statistically significant anomaly pattern of zonal wavenumber 2–3 is found around the edge of the polar vortex. The geopotential height at 500 hPa (not shown) is found to exhibit a response similar to that in SLP in both NH and SH.

The NH SLP anomaly pattern for MLS-analysis (Figure 4b) exhibits a negative Arctic Oscillation-like structure. This is consistent with Figure 3d [Wittman et al., 2005], where a weakening and equatorward shift of the jet is observed (dipole centered on the NH subtropical jet). Note also that the magnitude and significance of the anomaly pattern in Figure 4b are much stronger than those of MLS-LiShine (Figure 4a). In the SH, the MLS case underestimates (overestimates) the SLP at the pole (the polar vortex region just outside the Antarctic continent) and is consistent with the poleward shift and strengthening of the SH jet compared to the analysis (Figure 3d).

So far, we have demonstrated that there is some statistically significant impact on medium-extended range forecasts over the LiShine case when using MLS ozone data. However, these responses are generally small compared to the deviation from the analyses. In the following we examine to what extent medium-extended range forecasts will benefit from an improved representation of ozone. This is achieved by verifying the forecasted fields from each case against analyses using methods suggested in the manual by World Meteorological Organization [1992].

We now consider the evolution of the root-mean-square errors (RMSEs) (verified against analyses) of various fields. In the following, the SH, tropics, and NH are defined as the regions between 90°S and 20°S, 20°S and 20°N, and 20°N and 90°N, respectively. As discussed in section 2.2, there is no abrupt change in the anomaly pattern of zonal mean ozone in March and Figure 1b is representative of all March 2011 ensemble forecasts considered. The similarity in their RMSE responses shown in Figures 6 and 7 supports our decision of combining the results of all six ensemble forecasts for March 2011, effectively increasing the number of ensemble members from 24 to 144. Figure 5a shows the time series of RMSEs in air temperature at 10 hPa (T10). In the NH, there is no significant difference between the magnitude and development of the RMSEs of LiShine and MLS. In the tropics, the RMSEs of T10 of the experiments (Figure 5a, middle) begin to diverge on day 1 of the forecasts, and by day 16, the RMSE of the MLS case is 0.4 K lower than that of LiShine, which corresponds to an improvement of 16.1%. The differences in RMSE between the two experiments (both against analysis) reduce gradually toward the end of the forecasts. At 50 hPa (Figure 5b, middle), the RMSEs of LiShine and MLS do not diverge until day 20. By the end of the forecasts, the RMSEs of MLS is lower than those of LiShine by 10.0%. As the Earth approaches equinox, solar heating and the potential for photochemical reactions of ozone are maximum in the tropics. The apparently surprising temperature response near
the NH pole may be a result of dynamic effects being larger than the radiative and photochemical effects (which are relatively small compared to those at low latitudes). In the SH, improvements in stratospheric temperature are also observed in response to using MLS ozone.

Figures 5c and 5d show the RMSEs in horizontal wind vector at 10 hPa (V10) and 50 hPa (V50). There are no significant differences between the RMSEs of LiShine and MLS. In the tropics and SH, the RMSEs of the MLS case are generally lower than those of LiShine (with the exception of V10). However, the extent to which the 1σ error bars overlap shows that the improvement is not statistically significant.

Figure 6 shows the time series of RMSE (verified against analyses) of four tropospheric fields, namely horizontal winds at 250 hPa (V250) and 850 hPa (V850), geopotential height at 500 hPa (GH500), and SLP. The RMSEs in upper tropospheric winds of both LiShine and MLS evolve quickly over the first 10–12 days, irrespective of the ozone fields used in the forecast. Beyond day 12, the RMSEs hit an asymptote and stays roughly uniform till the end of the forecast. The only exception to this is V250 in the tropics (Figure 6a, middle), where the rate of growth of RMSE slows down but does not stop after day 12. Unlike the stratosphere, the RMSEs of tropospheric wind for MLS are virtually the same as LiShine, regardless of forecast range and
Figure 5. RMSEs (against analysis) of (a) temperature at 10 hPa, (b) temperature at 50 hPa, (c) horizontal wind vector at 10 hPa, and (d) horizontal wind vector at 50 hPa for the (top) NH, (middle) tropics, and (bottom) SH. RMSEs for the LiShine and MLS cases are shown in red and blue, respectively. Error bars show 1 standard deviation of the $24 \times 6$ ensemble members.
Figure 6. RMSEs (against analysis) of (a) horizontal wind vector at 250 hPa, (b) horizontal wind vector at 850 hPa, (c) geopotential height at 500 hPa, and (d) sea level pressure for (top) NH, (middle) tropics, and (bottom) SH. RMSEs for the LiShine and MLS cases are shown in red and blue, respectively. Error bars show 1 standard deviation of the 24 x 6 ensemble members.
Figure 7. Same as Figure 6 but for the ensemble forecast initialised on 1 March 2011 only. RMSEs for the LiShine, MLS, and MLSx2 cases are shown in red, blue, and green, respectively.

(a) V250 (m s$^{-1}$)  
(b) V850 (m s$^{-1}$)  
(c) GHS500 (m)  
(d) SLP (hPa)
region (NH, tropics, and SH) considered. The observed null result also applies to other verification fields (GH500 and SLP) in the troposphere, as shown in Figures 6c and 6d.

To verify the lack of sensitivity of RMSEs to the ozone field applied, a third set of ensemble forecasts, MLSx2, is run for 1 March 2011 case. The ozone field used in MLSx2 is structurally the same as MLS, except that its anomaly pattern (MLSx2-LiShine) is twice that of MLS-Shine (see Figure 1, right). Figure 7 shows that there is no significant difference between the RMSEs in MLSx2 and the other forecasts in any of the tropospheric fields considered. Given this result we did not consider it worthwhile investing in resources to investigate other starting dates. The same applies to the mean SLP maps (not shown) where the anomaly pattern is similar. This further confirms that the RMSEs of medium-extended range forecast made by MOGREPS are largely insensitive to the ozone field used.

4. Discussion and Conclusion

The aim of this paper was to investigate the impact of more realistic ozone data on medium to extended range forecasts made by the Met Office NWP system. This was achieved by performing 31 day forecasts using MOGREPS with monthly mean zonal mean ozone distributions computed from the MLS data set for the same periods; the quality of the MLS case was compared with standard MOGREPS runs which utilize a 5 year monthly mean zonal mean ozone field as described in Li and Shine [1995]. Several fields in the troposphere in each of these forecasts were verified against analyses using the method suggested by the World Meteorological Organization.

For a comprehensive study of the impact of using an alternative ozone field, experiments were repeated for several 1 month periods (see section 2.2). As the impact on tropospheric weather forecasts was found to be small and similar among these periods, we chose to present only results from the March 2011 case when an unusually large ozone depletion was observed over the Arctic in this paper. This “ozone hole” is not characterized by the Li and Shine data set which was based on earlier years. In addition, there is no abrupt change in ozone concentration over the period concerned. We consider this to be one of the best scenarios for assessment of the impact of using MLS ozone in the Met Office NWP system. A total of six ensemble forecasts were produced, which were identical in terms of the ozone fields used for the control and the experiment but differed in their dates of initialization.

Despite the fact that the MLS ozone data differ significantly from, and better describe the Arctic “ozone hole” in March 2011 than the Li and Shine climatology, results from the average of our six consecutive 24-member ensemble forecasts show that there is no significant improvement in the RMSEs (against analyses) of tropospheric fields at any forecast range considered. The zonal mean temperature and zonal wind responses of the Met Offices NWP system are, however, sensitive to ozone fields used. It is also interesting to note that the impact of the ozone change on the troposphere in this period is similar or larger than that seen in other experiments (not shown) run for northern and southern winter and southern spring. It takes at least 20 days for tropospheric zonal winds to fully respond to forcings in the stratosphere. However, at this forecast range, the deviation of forecast from analysis is dominated by the spread of ensemble members, rendering the benefits of adopting an accurate representation of monthly mean ozone negligible.

In the context of stratosphere-troposphere coupling, Simpson et al. [2009] demonstrated that the tropospheric temperature response exhibits a banded structure which developed starting from day 20–29 (less than 0.2 K) and continues to intensify until the model reaches equilibrium (0.25–0.5 K) in their ES experiment (stratospheric heating perturbation of 5 K at the equator reducing to 0 K at the poles). In our case, we observe a 1–1.5 K change at the equatorial lower stratosphere (reducing to less than 0.5 K near the poles) and a banded response of 0.2 K in the troposphere by day 21–31. Due to differences in methodology and model complexity [Simpson et al. [2009] used a simplified general circulation model with a direct thermal perturbation], it is difficult to directly relate the strength of the response in our study to theirs. We do observe, however, that the timescale (21–31 days) at which the banded response starts to develop in the troposphere is very similar to their findings.

Zonal asymmetries in ozone are often induced by planetary wave-driven displacement of the polar vortex [e.g., Ialongo et al., 2012]. The 2011 Arctic ozone hole showed considerable zonal asymmetry which was not uniform with time. It is possible that daily assimilation of a full 3-D ozone field could have more impact on the troposphere than our MLS runs, which rely on a prescribed zonal mean ozone distribution. For instance,
McCormack et al. [2011] reported that a 3-D prognostic ozone run would result in a warmer NH winter polar stratosphere compared to a zonal mean ozone run. However, they find that this is only significant at the 95% level between days 71 and 80, which is outside the forecast range considered in this paper. It is therefore unclear that 31 day forecasts with 3-D ozone fields would provide significant changes in tropospheric forecast fields. We were not able to test the impact of zonal asymmetry in this study, however, due to practical issues related to observation gaps.

In summary, our results show that even in the extreme case when one ozone field is clearly more appropriate than the other (i.e., March 2011 Arctic “ozone hole”), the choice of monthly mean zonal mean ozone climatologies has no significant improvement on medium-extended range forecasts made with current state-of-the-art NWP systems.

Appendix A: Statistical Significance

To assess the statistical significance of the results presented in this paper we make use of a Student’s t test, with adjustments to account for paired sample under serial dependence. We follow the method described in Wilks [2011], now summarized.

Let \( x_1 \) and \( x_2 \) be timeseries of equal length \( n \) of which the difference between their means is to be assessed. For a paired sample, the problem can be simplified to a 1-sample t test,

\[
\bar{\Delta} = \frac{1}{n} \sum_{i=1}^{n} \Delta_i = \bar{x}_1 - \bar{x}_2 \tag{A1}
\]

\[
z = \frac{\bar{\Delta}}{\left( \frac{\sum_{i=1}^{n} \Delta_i^2}{n} \right)^{1/2}} \tag{A2}
\]

where \( \bar{\Delta} \) is the mean difference of \( x_1 \) and \( x_2 \), \( z \) is the t statistics under null hypothesis, \( \sum_{i=1}^{n} \Delta_i^2 \) is the sample variance, and the overbar denotes that the quantity is time averaged. Note that \( x_1 \) and \( x_2 \) could refer to either analyses or ensemble mean of the forecasts.

To account for serial dependence in our data, we assume a lag-1 autocorrelation in \( \Delta \) and replace \( n \) in equation (A2) with an effective sample size \( n' \)

\[
n' \approx n \frac{1 - r_1}{1 + r_1} \tag{A3}
\]

The general expression for lag- \( k \) correlation coefficient is given by

\[
r_k = \frac{\sum_{i=1}^{n-k} \left( \Delta_i - \bar{\Delta}_- \right) \left( \Delta_{i+k} - \bar{\Delta}_+ \right)}{\left( \sum_{i=1}^{n-k} \left( \Delta_i - \bar{\Delta}_- \right)^2 \right)^{1/2} \left( \sum_{i=k+1}^{n} \left( \Delta_i - \bar{\Delta}_+ \right)^2 \right)^{1/2}} \tag{A4}
\]

where \( \bar{\Delta}_- \) and \( \bar{\Delta}_+ \) denote average of first and last \( n-k \) members of the difference series \( \Delta \), respectively. The benefit of the above treatment for paired and autocorrelated data is that most temporal dependence of both \( x_1 \) and \( x_2 \) is removed when calculating the differences. Also, \( r_1 \) in the difference series is generally smaller than that calculated from \( x_1 \) and \( x_2 \) individually; this results in a larger effective sample size and provides a more sensitive test [Wilks, 2011].

For convenience, each of the ensemble forecasts produced for March 2011 is named after their start dates so, for example, forecasts initialized on 1 March are referred to as EF1. To take account of the six ensemble forecasts in our assessment of statistical significance, we group data according to their forecast dates \( (t_i) \) (see section 3) and treat them as a single time series

\[
\{ EF1(t_1), \ldots, EF6(t_1), EF1(t_2), \ldots, EF6(t_2), \ldots, EF1(t_i), \ldots, EF6(t_i) \} \tag{A5}
\]

where we substitute \( k = 6 \) in equation (A4) such that the assumption of lag-1 autocorrelation within each ensemble forecast is still valid.
Acknowledgments
This project was funded by Natural Environment Research Council, in partnership with the Met Office. We thank Andrew Bushell for useful discussions and Andrew Charlton-Perez for commenting on an earlier draft of the paper. Data used in this paper are available upon request. Please contact david.jackson@metoffice.gov.uk for details.

References
Baldwin, M. P., and T. J. Dunkerton (2001), Stratospheric harbingers of anomalous weather regimes, Science, 294(5542), 581–584.
Bowler, N. E., A. Arribas, K. R. Mylne, K. B. Robertson, and S. E. Beare (2008), The MOGREPS short-range ensemble prediction system, Q. J. R. Meteorol. Soc., 134(632), 703–722. doi:10.1002/qj.234.
De Grandpré, J., R. Ménard, Y. J. Rochon, C. Charette, S. Chabrillat, and A. Robichaud (2009), Radiative impact of ozone on temperature predictability in a coupled chemistry-dynamics data assimilation system, Mon. Weather Rev., 137(2), 679–692.
Dethof, A., and E. V. Holm (2004), Ozone assimilation in the ERA-40 reanalysis project, Q. J. R. Meteorol. Soc., 130(603), 2851–2872.
Froyd, L., et al. (2006), Early validation analyses of atmospheric profiles from EOS MLS on the Aura satellite, IEEE Trans. Geosci. Remote Sens., 44(5), 1106–1121.
Haigh, J. D., M. Blackburn, and R. Day (2005), The response of tropospheric circulation to perturbations in lower-stratospheric temperature, J. Clim., 18(17), 3672–3685.
Ialongo, I., V. Sofieva, N. Kalloski, J. Tamminen, and E. Kyrölä (2012), Ozone zonal asymmetry and planetary waves characterization during Antarctic spring, Atmos. Chem. Phys., 12(5), 2603–2614.
Jung, T., M. J. Miller, and T. N. Palmer (2010), Diagnosing the origin of extended-range forecast errors, Mon. Weather Rev., 138, 2434–2446.
Li, D., and K. Shine (1995), A 4-dimensional ozone climatology for UGAMP models, UGAMP Internal Rep. 35, 35 pp., Center for Global and Atmospheric Modelling, Department of Meteorology, Univ. of Reading, U. K.
Manney, G. L., et al. (2011), Unprecedented Arctic ozone loss in 2011, Nature, 478(7370), 469–475. doi:10.1038/nature10556.
Mathison, C., D. R. Jackson, and M. Keil (2007), Methods of improving the representation of ozone in the Met Office model, NWP Tech. Rep. 502, Met Office, Exeter, U. K.
McCormack, J. P., T. R. Nathan, and E. C. Cordero (2011), The effect of zonally asymmetric ozone heating on the Northern Hemisphere winter polar stratosphere, Geophys. Res. Lett., 38, L03802, doi:10.1029/2010GL045937.
Morcrette, J. J. (2003), Ozone-Radiation Interactions in the ECMWF Forecast System, Technical Memorandum (European Centre for Medium Range Weather Forecasts), vol. 375, European Centre for Medium-Range Weather Forecasts, Reading, England. U. K.
Ndarnana, T., D. W. Waugh, L. M. Polvani, G. J. P. Correa, and E. P. Gerber (2012), Antarctic ozone depletion and trends in tropopause Rossby wave breaking, Atmos. Sci. Lett., 13(3), 164–168.
Orr, A., T. J. Bracegirdle, J. S. Hosking, W. Feng, H. K. Roscoe, and J. D. Haigh (2012), Strong dynamical modulation of the cooling of the polar stratosphere associated with the Antarctic ozone hole, J. Clim., 26(2), 662–688.
Polvani, L. M., and J. J. Kusner (2002), Tropospheric response to stratospheric perturbations in a relatively simple general circulation model, Geophys. Res. Lett., 29, 1114, doi:10.1029/2001GL014284.
Randel, W. J., and F. Wu (1999), A stratospheric ozone trends data set for global modelling studies, Geophys. Res. Lett., 26, 3089–3092.
Rishigjaard, L. P. (1996), On four-dimensional variational assimilation of ozone data in weather-prediction models, Q. J. R. Meteorol. Soc., 122(535), 1545–1571.
Semane, N., V.-H. Peuch, S. Pradier, G. Desroziers, L. El Amraoui, P. Brousseau, S. Massart, B. Chapnik, and A. Peuch (2009), On the extraction of wind information from the assimilation of ozone profiles in Météo France 4-D-Var operational NWP suite, Atmos. Chem. Phys., 9, 4855–4867.
Simpson, I. R., M. Blackburn, and J. D. Haigh (2009), The role of eddies in driving the tropospheric response to stratospheric heating perturbations, J. Atmos. Sci., 66(5), 1347–1365.
Stajner, I., et al. (2008), Assimilated ozone from EOS-Aura: Evaluation of the tropopause region and tropospheric columns, J. Geophys. Res., 113, D16532, doi:10.1029/2007JD008863.
Wilkus, D. S. (2011), Statistical Methods in the Atmospheric Sciences, vol. 100, Academic Press, Oxford, U. K.
Williams, G. P. (2006), Circulation sensitivity to tropopause height, J. Atmos. Sci., 63(7), 1954–1961.
Wittman, M. A. H., A. J. Charlton, and L. M. Polvani (2005), On the meridional structure of annular modes, J. Clim., 18, 2119–2122.
World Meteorological Organization (1992), Manual on the Global Data-Processing System, Secretariat of the World Meteorological Organization, Geneva, Switzerland.