

887

Evaluating the impact of prognostic ozone in IFS NWP forecasts

R. S. Williams, R. J. Hogan, I. Polichtchouk,
M. I. Hegglin (U. Reading), T. Stockdale and
J. Flemming

October 2021

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:

<http://www.ecmwf.int/publications/>

Contact: library@ecmwf.int

© Copyright 2021

European Centre for Medium Range Weather Forecasts
Shinfield Park, Reading, Berkshire RG2 9AX, England

Literary and scientific copyrights belong to ECMWF and are reserved in all countries. The content of this document is available for use under a Creative Commons Attribution 4.0 International Public License. See the terms at <https://creativecommons.org/licenses/by/4.0/>.

The information within this publication is given in good faith and considered to be true, but ECMWF accepts no liability for error, omission and for loss or damage arising from its use.

Abstract

The forecast impact of a recently developed, novel prognostic ozone scheme, interactive with radiation, is here evaluated within the ECMWF Integrated Forecasting System (IFS) on both medium-range and sub-seasonal timescales. The importance of radiatively interactive ozone has been demonstrated in terms of accurate representation of stratospheric heating rates, which has direct implications for the predictability of temperature and wind. At ECMWF, the capacity to implement prognostic ozone interactive with radiation, on an operational forecast basis, has existed within IFS for several years. However, such usage has been inhibited by insufficient signal for an overall skill improvement. The new Hybrid Linear Ozone (HLO) scheme, evaluated here, simulates the impact of chemistry on the tendency of ozone, using a blend of inputs from analyses and chemistry model calculations, in a realistic and affordable manner. The scheme has been developed to address a number of shortcomings associated with earlier linear ozone schemes, such as the widely used Cariolle parametrization. With the aid of different established verification metrics, the forecast skill impact of implementing radiatively interactive HLO is critically evaluated, in regard to both changes to the variability aspect of skill (e.g. pattern correlations) and the mean bias. A mixture of individual and aggregate (e.g. multi-month) cases are examined, using a combination of deterministic and ensemble forecasts (comparing interactive and control experiments). Large changes in stratospheric ozone in association with midwinter Sudden Stratospheric Warmings (SSWs) additionally provide natural candidates for verification of the HLO scheme performance, particularly as such events are known to impact tropospheric weather, and are thus a key focus of this report. We find a clear overall improvement in the representation of forecast temperature, wind and geopotential height in 15-day forecasts in the stratosphere, compared to using an ozone climatology, supporting the inclusion of radiatively interactive HLO within the IFS medium-range forecasting system. In contrast, implementation of HLO, interactive with radiation, over longer sub-seasonal timescales does not appear to improve sub-seasonal forecasts following SSWs. A key limitation of this focus however is the already enhanced forecast skill following these events, as the stratospheric state is inherently more predictable, and the performance of radiatively interactive HLO during more typical wintertime conditions has not here been assessed. Nevertheless, radiatively interactive HLO is demonstrated to yield enhanced lower stratospheric forecast temperature skill following SSWs, in conjunction with the persistence of a long-lived positive anomaly in ozone. Detailed investigation of the evolution in radiative heating tendencies following the major January 2009 SSW provides insight into the mechanisms of the response to radiatively interactive HLO and the reasons why such implementation does not necessarily translate to forecast skill improvement. The findings are used to highlight a series of recommendations for further verification and refinement of the HLO scheme.

1. Introduction

An important limitation of numerical weather prediction (NWP) systems is that most include only a simplistic representation of atmospheric chemistry. Atmospheric trace gases such as ozone (O_3) and water vapour (H_2O), which are profoundly variable in space and time, have an important radiative influence on timescales ranging from days to weeks (weather) up to several decades (climate). The globally averaged radiative forcing, due to anthropogenic changes to ozone since preindustrial times, was estimated to be $+0.35$ (0.15 to 0.55) $W\ m^{-2}$ in the Intergovernmental Panel for Climate Change Fifth Assessment Report (IPCC AR5), with a contribution of $+0.40$ (0.20 to 0.60) $W\ m^{-2}$ from the troposphere and -0.05 (± 0.10) $W\ m^{-2}$ from the stratosphere. Unchanged from the IPCC AR4 report, the radiative forcing due to stratospheric water vapour produced by methane (CH_4) oxidation was estimated to be $+0.07$ (± 0.05) $W\ m^{-2}$ (Hansen et al., 2005; Myhre et al., 2013). Because the spatiotemporal distribution of ozone and water vapour is highly variable however, the radiative forcing exerted by these chemical constituents varies significantly with latitude (e.g. Riese et al., 2012; Checa-Garcia et al., 2018). On much shorter timescales ranging from days to months, the spatiotemporal evolution of stratospheric ozone and water vapour can have important transient radiative impacts which are pertinent to the accurate prediction of stratosphere-troposphere coupling processes (e.g. Shepherd, 2002; Haase and Matthes, 2019). Such mechanisms in which this may impact predictability of synoptic-scale features include modification of: large-scale horizontal shortwave atmospheric heating gradients (which may perturb mean winds through thermal-wind balance); the vertical heating profile (which has implications for tropospheric stability and hence convection); and atmospheric heating (which may act as a Rossby-wave source with impacts on the global circulation; Rodwell and Jung, 2008). Impacts on tropospheric forecast skill can therefore manifest on both medium-range forecast timescales (typically up to 10-15 days) and on longer sub-seasonal to seasonal timescales.

The radiative effect of both ozone and water vapour is known to be significant in the upper troposphere-lower stratosphere (UTLS) region (e.g. Randel and Wu, 2010; Gilford et al., 2016), and changes in such radiatively active trace gas species in this region are known to have the largest relative impact on surface temperatures (Lacis et al., 1990; Riese et al., 2012). Indeed, pronounced surface climate impacts across the southern hemisphere have been unambiguously detected in recent decades, as a consequence of both radiative and dynamical effects of the Antarctic ozone hole (e.g. Thompson et al., 2012; Kidston et al., 2015; Byrne et al., 2019). Such changes largely project onto the southern annular mode (SAM), with the most robust impacts occurring in austral summer (Thompson and Solomon, 2002), in conjunction with a poleward displacement of the tropospheric jet (Arblaster and Meehl, 2006). This leads to modification of air-sea fluxes through perturbed wind stresses (Cagnazzo et al., 2013) with a possible,

but unclear, causal link to carbon ocean uptake (Fahey et al., 2018) and Atlantic thermohaline circulation behaviour (e.g. Reichler et al., 2012). The implementation of interactive ozone in models may furthermore have implications for climate change projections. For instance, Dietmüller et al. (2014) found that the climate sensitivity parameter is dampened by 8.4%, due to a negative stratospheric ozone (and associated change in stratospheric water vapour) feedback, in response to a quadrupling of CO₂ using the EMAC chemistry-climate model (CCM). Nowack et al. (2015) further found indication of a reduction in surface warming due to 4 x CO₂ (up to 20%), resulting from changes to the Brewer-Dobson Circulation (BDC) and subsequent feedbacks to the chemical composition of the lower stratosphere. On the contrary, other studies which have performed similar experiments conclude that feedbacks due to stratospheric ozone chemistry are not of significant importance in determining climate sensitivity (e.g. Marsh et al., 2016). A later study (Nowack et al., 2017) also serves to counter the model consensus of a likely increase in El Niño conditions under climate change, resulting from an induced negative feedback to projected Walker Circulation (and to some extent Hadley Circulation) changes.

The importance of interactive chemistry for accurate model representation of stratosphere-troposphere coupling on intraseasonal timescales has been further demonstrated in the context of stratospheric extreme events, which namely includes sudden stratospheric warmings (SSWs) (Haase and Matthes, 2019; Oehrlein et al., 2020). Further investigation is however required in understanding the benefits this may yield, particularly in forecasting systems on both medium-range and sub-seasonal to seasonal timescales, which is a focus of this investigation within the IFS. The implications of anomalously large changes in stratospheric chemical composition (specifically ozone) for NWP are visited next in section 1.1 during SSWs.

1.1. Sudden stratospheric warmings (SSWs)

SSWs comprise the largest deviations from the mean state on the wintertime northern hemisphere extratropical stratosphere, with a frequency of ~6 events per decade (Charlton et al., 2007), and are defined by a reversal of the climatological polar temperature gradient and the 60°N circumpolar mean wind from westerly to easterly at 10 hPa (~30 km) (Andrews et al., 1987). SSWs strongly impact the distribution of chemical constituents within the stratosphere (e.g. Manney et al., 2009a; 2009b; Tao et al., 2015), but the radiative implications of these changes in the radiatively sensitive UTLS have not been extensively studied. As reported by Manney et al. (2009a) and Tao et al. (2015), typical changes around the SSW onset include enhanced meridional transport and polar descent in the stratosphere, facilitated by an acceleration of the stratospheric BDC. The result is an increase in stratospheric ozone, as well as a breakdown in the polar vortex mixing barrier, that leads to a flattening of tracer gradients in species such as water vapour (H₂O), carbon monoxide (CO), methane (CH₄) and nitrogen dioxide (N₂O), due to enhanced mixing across the weakened polar vortex edge. In terms of longitudinal

structure, many filamentary structures in ozone manifest in conjunction with such events over a broad range of spatial scales (e.g. McIntyre and Palmer, 1983; Konopka et al., 2003; Grooß et al., 2005), whilst the distribution of stratospheric ozone is largely zonally symmetric when the stratospheric polar vortex (SPV) is strong (enhanced zonal flow). Thus, NWP systems, which typically only include a simplistic representation of stratospheric ozone (e.g. a zonal-mean, monthly-mean climatology) in the radiation scheme, are likely to be more deficient in accurately simulating the coupling between dynamics, transport and chemistry during the evolution of such extreme events.

The implementation of radiatively interactive, 3D prognostic ozone fields within NWP systems could therefore lead to a substantial increase in forecast skill around these times, when the distribution and abundance of ozone deviates strongly from climatology. The significance of this is enhanced further by the larger influence of the stratosphere on tropospheric weather regimes following SSWs, which is of particular interest to forecast centres such as ECMWF. On the other hand, the already enhanced stratospheric and tropospheric predictability following SSW events (e.g. Sigmond et al., 2013) perhaps limits further skill improvements prognostic ozone interactive with radiation may provide. This is exemplified in Figure 1 using the temperature anomaly correlation coefficient (ACC) skill metric, computed as an average over the polar-cap region for each hemisphere in the 60 days following the January 2009 SSW, for the ensemble-mean of a control forecast experiment (initialised using the IFS on the central warming or onset date). Values greater than 0.6 denote a high level of skilful predictability, which can be seen to persist in the Arctic stratosphere long after the event (40-50 days), in sharp contrast with the summertime Antarctic stratosphere (~10 days) and indeed elsewhere in the stratosphere at this time (not shown). Nevertheless, midwinter SSW events still present an ideal

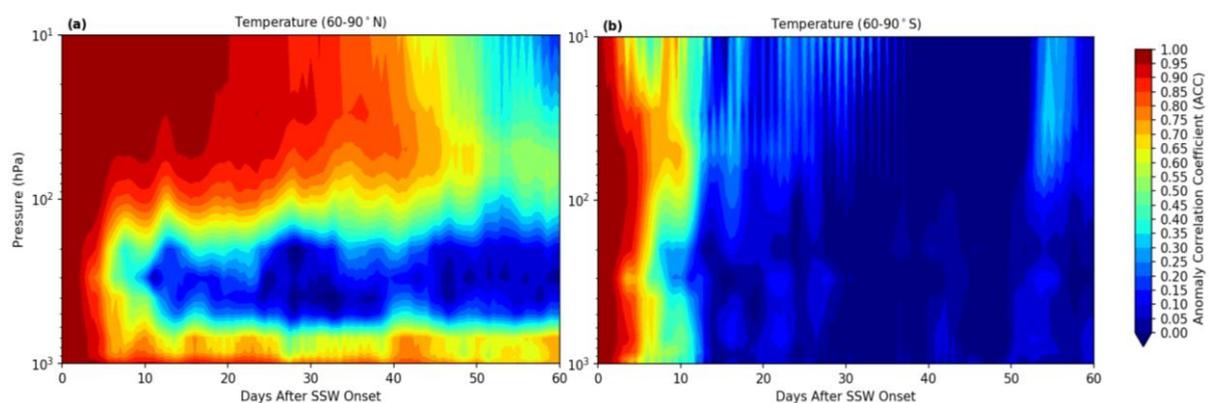


Figure 1: The polar-cap averaged evolution in the temperature anomaly correlation coefficient (ACC) for (a) the northern hemisphere (60-90°N) and (b) the southern hemisphere (60-90°S), as a function of pressure (10-1000 hPa) and time following the central warming (onset) date of the 24th of January 2009 northern hemisphere SSW event. This result is derived from a control forecast experiment performed using a sub-seasonal configuration in the IFS (see section 2.2 for details).

opportunity to investigate the potential skill gain that such schemes may offer, including the mechanisms involved, and are thus subject of scrutiny in this report.

An SSW can be dynamically distinguished as either a vortex displacement or vortex split event as illustrated in Figure 2, which is typically forced by a primary wave-1 or wave-2 perturbation to the zonal westerly flow respectively. An alternative distinction is made based on the two leading empirical orthogonal functions (EOFs) of daily-mean polar-cap (70-90°N) temperature anomalies integrated over the middle atmosphere column, with so-called Polar-night Jet Oscillation (PJO) events identified where the SSW temperature anomaly (as projected onto these two EOFs) is maximised at ~60 hPa, so long as the anomaly is sufficiently strong (Kuroda and Kodera, 2004; Hitchcock et al., 2013). Such distinction is related to the depth to which the warming descends through the stratosphere, which is closely associated with the magnitude of upward and poleward directed wave forcing. PJO events are noted for the subsequent strength of their signal in the lower stratosphere. Using ERA-Interim reanalysis and Whole Atmosphere Community Climate Model (WACCM) simulations, de la Cámara et al. (2018b) showed that a weaker residual circulation and enhanced isentropic mixing establishes for up to three months in the lower stratosphere, with a much more muted and shorter-lived response following non-PJO events. The weaker residual circulation is a feature of the recovery stage of the SPV, following a strengthened residual circulation leading up to the SSW onset. PJO-composite anomalies in isentropic mixing were found to be almost twice as large throughout the stratosphere, whereas any enhanced mixing within the lower stratosphere during non-PJO events was determined to be largely statistically insignificant in the de la Cámara et al. (2018b) study.

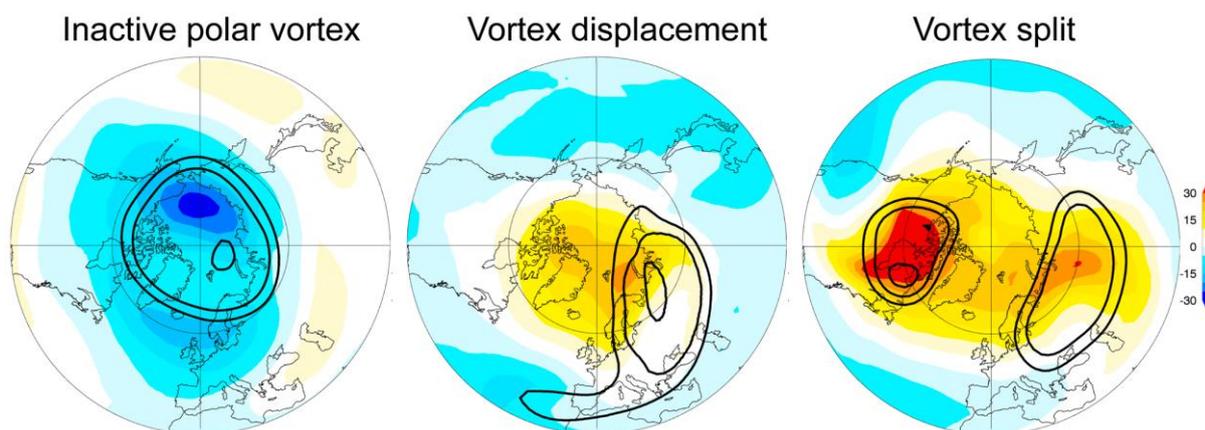


Figure 2: A depiction of a vortex displacement (23 January 1987) (centre) and vortex split event (24 January 2009) (right) in relation to an inactive (strong) state of the wintertime stratospheric polar vortex (~9 January 2009) (left) from the MERRA-2 reanalysis. Contours denote potential vorticity at 550 K (for 75, 100 and 125 PVU) and shading represents temperature anomalies (K) at 10 hPa. Taken from Figure 1 of Butler et al. (2017).

The PJO subclass of events therefore have a stronger impact on the composition of the lower stratosphere, given deeper propagation down to the lowermost stratosphere (LMS) (tropopause to 100 hPa) and known dynamical persistence timescales of up to 2-3 months (Hitchcock et al., 2013; de la Cámara et al., 2018a). The tropospheric circulation is thus likely modulated by SSWs that propagate down to the tropopause through stratosphere-troposphere coupling. Although the mechanisms surrounding stratosphere-troposphere coupling following SSWs are poorly understood, likely involving eddy feedbacks in the troposphere (e.g. Domeisen et al., 2013), it is nevertheless clear that the build-up and onset of such events provides a significant source of predictability on sub-seasonal to seasonal timescales.

1.2. Stratospheric ozone schemes

Unlike CCMs and chemical transport models (CTMs), which are specifically tailored to address understanding of the interplay between chemistry and dynamics that govern the formation, destruction and resultant distribution of stratospheric ozone, comprehensive chemistry schemes are not feasible for implementation in operational NWP models due to the high computational cost (Monge-Sanz et al., 2021). Therefore, realistic and affordable chemistry schemes are required to balance this constraint with the potential forecast skill gain that may be yielded. Whilst a reasonable representation of water vapour (interactive with radiation) exists in NWP models, notwithstanding some issues in the upper troposphere-lower stratosphere (UTLS) region related to modelling deficiencies (e.g. numerical diffusion) and the absence of any observational assimilation above the tropopause, representation of radiatively interactive ozone is often much more simplistic as earlier discussed.

Common to many operational systems, the radiation scheme within the ECMWF IFS model uses a zonal-mean, monthly-mean ozone climatology as a basic description of latitudinally varying stratospheric ozone abundance and distribution. This climatology is derived from the Copernicus Atmospheric Monitoring Service (CAMS) reanalysis (as of IFS Cycle 43r1; implemented in November 2016), which assimilates observations from various different satellite sources, including the Microwave Limb Sounder (MLS), the Solar Backscatter Ultraviolet and Total Ozone Monitoring Spectrometer (SBUV) and the Global Ozone Monitoring Experiment-2 (GOME-2) instrument (Inness et al., 2019). Stratospheric ozone chemistry within CAMS is parametrized using the first linear ozone model of stratospheric ozone formulated by Cariolle and Déqué (1986), hereafter referred to as the CD scheme. Although the capability of using prognostic ozone (interactive with radiation) has existed in the IFS for some time, the only longstanding implementation operationally to date was in the System 4 seasonal forecasting system (using the CD scheme), where improvements in the predictability of the QBO were demonstrated (Hogan et al., 2017). Interactive prognostic ozone was not used in the subsequent SEAS5 seasonal forecasting system.

The temporal evolution of ozone concentration within the default ECMWF (CD) scheme is specified by the following equation:

$$\frac{\partial f}{\partial t} = c_0 + c_1(f - \bar{f}) + c_2(T - \bar{T}) + c_3(c_{O_3} - c_{\bar{O}_3}) + c_4(Cl_{EQ})^2 f \quad (1)$$

where f is the ozone concentration, T is the temperature, c_{O_3} the overlying partial ozone column and Cl_{EQ} the chlorine content equivalent in the stratosphere (which is variable from year to year). All terms with overbars constitute climatological reference values (obtained from CAMS full-chemistry output fields). Coefficient c_0 constitutes the climatological mean production rate of ozone. Gas-phase chemical effects are specified by the tendency coefficients c_i ($i = 0, 1, 2, 3$) with the additional coefficient c_4 operationally available to account for ozone destruction related to heterogeneous chemical processes at very low temperatures (<195 K during daytime poleward of 45° latitude) (Cariolle and Teyss re, 2007). The effective chlorine content parameter (Cl_{EQ}) uses an annual time series of values based on model prediction of the ozone depleting substances (ODS) burden. A recommended application of a cold tracer to track temperatures below the formation threshold of polar stratospheric clouds (PSCs) is under development but not yet operationally implemented within the IFS. The CD scheme parametrizes the effects of ozone gas-phase chemistry to simulate the evolution of stratospheric ozone with reasonable accuracy over broader spatiotemporal scales, as derived from the MOBIDIC 2D photochemical model (Cariolle and Teyss re, 2007). However, the CD scheme is not effective in simulating complex variability on day-to-day and/or synoptic spatial scales. The impact of heterogeneous chemical processes, such as is responsible for the development of the austral spring Antarctic ozone hole, is also not accounted for consistently (different methods and approximations for the c_4 term are used with respect to the other coefficients in Equation 1).

Over the last ten years, a new interactive linear ozone model has been tested and developed, with implementation capability within the ECMWF system (the BMS scheme; Monge-Sanz et al., 2011). The scheme can provide prognostic ozone in both medium-range and long-range forecast experiments. In experiments with prognostic ozone interactive with radiation, a clear indication of an improvement in the stratospheric mean temperature bias, as well as the seasonal evolution of the southern hemisphere SPV has been found for the BMS scheme (Monge-Sanz et al., 2021). The BMS scheme, like the CD scheme, consists of a linear representation of stratospheric ozone sources and sinks, determined as a function of ozone concentration and temperature. Dissimilar to the CD scheme or any other existing linear ozone model however, the BMS treats both gas-phase and heterogeneous stratospheric ozone chemistry consistently to ensure more realistic interactions between parametrized ozone, radiation and temperature to yield an improved response and feedback simulation to meteorological conditions compared with earlier parametrization schemes, particularly at high latitudes (Monge-Sanz et al., 2021). This is achieved through derivation of a set of coefficients from the 3D TOMCAT/SLIMCAT full-

chemistry CTM (Chipperfield, 2006), which includes all heterogeneous processes. As mentioned, prognostic ozone using the CD scheme was able to provide some benefit in the System 4 seasonal forecasting system, particularly concerning representation of the QBO and southern hemisphere springtime SPV variability, with similar results replicated when substituting in use of the BMS scheme. However, prognostic ozone was not implemented in the later SEAS5 seasonal prediction model, monthly or medium-range systems, following a slightly negative forecast skill impact on tropospheric variables (Hogan et al., 2017).

As an alternative effort to remedy the limitations of using the CD scheme for parametrization of stratospheric ozone chemistry, a prognostic hybrid linear ozone (HLO) scheme has been developed and is available for testing within the IFS. It is anticipated that this will be included as part of the next major IFS cycle (Cycle 48r1), together with a new and improved solar spectrum and semi-Lagrangian vertical filter (SLVF) (Polichtchouk et al., 2021). Again, the scheme entails a broadly similar linearisation approach to the default CD scheme to represent the chemical impact on the ozone tendency:

$$\frac{\partial f}{\partial t} = c_0 + c_1(f - \bar{f}) + c_2(T - \bar{T}) + c_3(c_{O_3} - c_{\bar{O}_3}) \quad (2)$$

Whereas the four coefficients are extracted from a 2D chemistry model in the CD scheme, the HLO scheme acquires a blended input from either analyses (temperature and wind from ERA-5 and both climatological mean ozone and overhead ozone column values from CAMS) or adapted chemistry model calculations (rate sensitivity estimates for ozone, temperature and overhead column ozone, modified to ensure internal consistency, which is particularly important in regions influenced by heterogeneous chemistry). Although the HLO scheme is deficient in fully reproducing the impact of heterogeneous chemistry (only the climatological effect is accounted for), the scheme is a significant advance on the CD scheme in that a 3D ozone field that is consistent with CAMS reanalyses, with respect to the mean climatology, can be derived within the IFS (Polichtchouk et al., 2021). The realism of simulated fields in ozone is, however, more dependent on the observations assimilated, the data assimilation method and background errors (i.e. quality of the analysis), at least in the medium range. Although it has been verified that ozone analyses derived using the HLO scheme within the IFS are slightly improved relative to the CD scheme from CAMS experimentation, with the exception of ozone hole conditions, it is stressed that such differences are small compared with differences that may arise from the choice of linear chemistry used after longer lead times (e.g. sub-seasonal range) or in model runs without ozone data assimilation.

The HLO scheme has been found to perform well with respect to both interannual and synoptic variability, with c_0 (the climatological mean production rate) constituting the key term in Equation 2 as it is diagnosed from an ozone field nudged towards CAMS analyses using a multi-year IFS run. The main advantage of the HLO scheme over both the default CD and more recent BMS scheme, is the

reduction in drift (development of systematic biases as the forecast evolves). The improved mean state can be largely attributed to the constraint of CAMS reanalysis fields (as opposed to full chemistry models), including the mean production rate of ozone. The CD scheme especially is characterised by large biases in the mean, which are significantly reduced in the HLO scheme. This enables the benefit of radiatively interactive ozone to emerge in forecast skill metrics, without degradation over longer timescales due to large drifts in the mean. Since IFS Cycle 47r1 (October 2020), radiatively interactive prognostic HLO has been implemented operationally within the CAMS system (Copernicus, 2019).

1.3. Precedent of interactive prognostic ozone in models

Most NWP systems run by leading forecast operational centres (e.g. ECMWF, NCEP and UKMO) do not currently implement radiatively interactive prognostic ozone. The primary constraint is the computational cost of simulating 3D ozone fields, involving repeat assimilation of observations from multiple sources and projection using complex chemistry model calculations for every forecast initialisation. To overcome this issue, schemes are being developed to balance the constraint of computational cost with the skill benefit such implementation may yield. Verification results of such schemes in different NWP models, often performed on an ad-hoc or test basis, are increasingly becoming more accessible. A short review of some recent findings is provided here.

To first highlight earlier work characterising the importance of interactive ozone chemistry, Sassi et al. (2005) showed that implementation of interactive ozone chemistry results in significant impacts on the climate of the middle atmosphere (via modulation to SW heating rates in particular) within the NCAR Whole Atmosphere Chemistry Climate Model (WACCM), with respect to a simulation using a constructed zonal-mean monthly-mean climatology. Most notably, a reduction in SW heating by 2K day⁻¹ (up to 50% of the total SW heating) is induced in the lower mesosphere (~65 km), as the effects of the diurnal cycle in ozone at this altitude are better captured, which in turn leads to an enhancement of ~0.2 K day⁻¹ in the upper stratosphere (~45 km) as more SW absorption of photons can occur here. Both features induce a large-scale statistically significant impact on temperature in all seasons, except over the wintertime polar region. McCormack et al. (2006) investigated the performance of a new CHEM2D-Ozone Photochemistry Parametrization (CHEM2D-OPP), implemented within the prototype high-altitude NOGAPS-ALPHA NWP forecast model for short-term (6 day) and long-term (1 year) simulations of stratospheric ozone. As verified against both individual Aura MLS measurements and independent ozone analysis fields (for the northern hemisphere), this scheme resulted in the best overall agreement, in terms of mean forecast error and RMSE, when compared with other ozone photochemistry parametrization schemes that were currently in use within operational NWP systems (e.g. the CD scheme at ECMWF) at the time of publication. The work of Jeong et al. (2016) corroborates such finding in that inclusion of radiatively interactive prognostic ozone fields, as simulated using linear

photochemistry parametrization schemes, reduced forecast error in ozone (validated again with respect to MLS and ozonesonde observations) within the Korean Institute of Atmospheric Prediction Systems (KIAPS) global forecast model. However, these studies did not quantify the forecast impact on key meteorological variables such as temperature and wind.

A study by De Grandpré et al. (2009) highlighted the importance of ozone data assimilation for systematic and large-scale improvements in projected temperature, with respect to radiosonde and Michelson Interferometer for Passive Atmospheric Sounder (MIPAS) observations, using the Canadian Global Environmental Multiscale (GEM) model, in conjunction with the Belgian Atmospheric Chemistry Module (BACH). When data assimilation is performed, predictive skill in temperature was found to improve by more than a day at 50 hPa across the northern hemisphere during summertime for instance, with a concurrent reduction in RMSE and bias over the assimilation period. Such benefits were not however realised in the case of no ozone data assimilation (i.e. use of a monthly mean climatology), when compared with an equivalent non-radiatively interactive simulation. A later study by Cheung et al. (2014) compared the impact of including a more realistic representation of ozone in the Met Office NWP system, from MLS measurements, with respect to operational usage of a coarse-resolution (2.5°) zonal-mean monthly mean climatology over a 5-year period (from Li and Shine, 1995), during a period of anomalously low Arctic ozone during March 2011. Whilst the zonal-mean temperature and wind responded to the choice of ozone field used, no clear overall improvement was found in the representation of tropospheric fields at any forecast range (up to a month lead time); consistent with earlier findings (e.g. Mathison et al., 2007). However, such investigation has not been extended to data assimilation of a full 3D ozone field which may have yielded a positive impact.

Coopmann et al. (2020) described the mechanisms by which comprehensive representation of atmospheric trace gas profiles may yield improved weather forecasts, through data assimilation at the point of forecast initialisation. In particular, the spatial and temporal distribution of ozone influences the spectral radiances received by hyperspectral infrared sounders. Extraction of this information can help minimise biases in ozone-sensitive channels and others (e.g. CO_2), as investigated using the ARPEGE model, which heralds improvement in analyses and the agreement of short-term forecasts (or analyses) to observations (e.g. radiosondes, microwave radiances). A significant improvement in short-term forecast error is yielded ($\sim 2.5\%$ for temperature, 4% for relative humidity and $\sim 3.5\%$ for wind over a wide expanse of atmosphere across a range of forecast lead times), by substituting a single fixed vertical profile in ozone with realistic ozone profiles simulated by the MOCAGE CTM (within the operational operator of the radiative transfer model used in ARPEGE). Such findings build upon earlier work which found that 4D assimilation of ozone reduced the lower stratospheric wind bias in ARPEGE (Semane et al., 2009). However, development of a prognostic ozone scheme is still under development

and so the impact of radiative feedbacks upon dynamics have not yet quantified using this model, which would likely further augment such skill improvement.

The forecast skill improvement radiatively active 3D prognostic ozone may yield in NWP forecasting systems can therefore be attributed to the following main aspects: (1) improved accuracy of operational radiance assimilation (e.g. Derber and Wu, 1998; John and Buehler, 2004); (2) improved accuracy of stratospheric radiative heating rates (e.g. Jackson and Saunders, 2002); and (3) improved representation of chemical, radiative and dynamical feedbacks, increasing forecast accuracy of stratospheric temperature and wind (e.g. Peuch et al., 2000). Earlier work synthesised here serves to highlight the novelty of this report in understanding the radiative feedbacks of both assimilating and projecting realistic 3D ozone fields as the forecast evolves.

1.4. Report structure

In this report, preliminary results of the impact of the radiatively interactive prognostic HLO scheme (implemented within the IFS) are presented and discussed, in anticipation that such a scheme could soon supersede the existing 2D ozone climatology. Recommendations for further verification of the scheme is provided, together with suggested actions for improvement. The evaluations herein were performed with the aim of understanding how prognostic ozone, interactive with radiation, affects forecast skill over a range of timescales from days to several weeks (up to 60 days). This includes not only quantification of the change in mean forecast error, but also involving different skill metrics. Another aspect is to elucidate and understand the mechanisms by which radiatively active ozone influences temperature and wind over different timescales (medium range and sub-seasonal). The below research questions were outlined in steering this research endeavour:

- 1) How does radiatively active prognostic ozone, using the HLO scheme, influence forecast skill in the IFS on medium-range (up to 15 day) and sub-seasonal (up to 60 day) timescales? How does the impact compare between each hemisphere (e.g. summer versus winter)?
- 2) How do projected spatial anomalies in ozone affect forecast temperature? Through thermal-wind balance, what impact does this subsequently have on spatial wind patterns?
- 3) Centred on northern hemisphere midwinter sudden stratospheric warming (SSW) events, how does prognostic HLO, interactive with radiation, impact forecast skill following such events? Is the dynamical coupling of stratospheric skill into the troposphere positively impacted by such implementation?
- 4) What are the radiative mechanisms by which the typical polar-cap evolution of ozone following PJO-type SSWs affects the forecast response? How does this impact translate to skill change spatially (both vertically and horizontally)?

This report is structured as follows: a description of the IFS forecast experiments performed and other ancillary datasets is provided in section 2, in addition to an overview of the verification metrics used in evaluating the performance of the HLO scheme (interactive with radiation). In section 3, results of the implementation of radiatively interactive HLO ozone are discussed in medium-range forecasts (up to 15 days). As well as evaluating the global impact on forecast skill over multi-month periods, the radiatively interactive HLO performance during a northern hemisphere SSW in February 2018 is additionally presented, in anticipation that the forecast evolution following the event may be more skilfully predicted. To complement the findings of this section, the following section 4 explores the performance of the new HLO ozone scheme within sub-seasonal forecasts (up to 60 days), centred on such extreme stratospheric events. Multiple PJO-type SSW events between 2000 and 2020 have been identified for analysis, with the statistical robustness of the results enhanced by initialisation of experiments (centred on the SSW central warming dates) with a full 51-member ensemble set. To elucidate and understand more clearly the radiative impact of ozone on temperature, a more in-depth investigation is undertaken for the pronounced SPV split event during the January 2009 SSW event, in which radiative heating tendencies due to ozone are extracted directly from the model forecasts. The results of this investigation are shown in section 5. The findings of this report are collectively synthesised in section 6, where a list of recommendations is provided to inform future development of the HLO scheme within the IFS on both medium-range and long-range (sub-seasonal) time frames. Where further evaluation of the HLO scheme performance is warranted, a summary of possible next steps is provided to guide this work. Finally, the main conclusions of this report are outlined in section 7.

2. Data and Methods

To quantify the impact of including the radiatively interactive prognostic HLO scheme within IFS NWP forecasts, a series of experiments have been performed over both medium-range (0-15 day) and long-range/sub-seasonal (0-60 day) timescales, initialised for different start dates in recent years. For each initialisation date, a pair of experiments have been run that includes both a control (in which a zonal-mean monthly-mean climatology is used in the radiation scheme, albeit revised from the currently operational ECMWF ozone climatology towards that applicable to the HLO scheme) and a radiatively interactive forecast (using prognostic 3D ozone fields parametrized using the HLO scheme). Note however that it is not only the HLO scheme which determines the distribution of ozone, as this is also influenced by non-linear advection by the dynamical core. Whilst the medium-range experiments performed included only one integration (i.e. deterministic forecasts), all sub-seasonal experiments consisted of a total of 51 ensemble members (i.e. probabilistic forecasts). A complete list of the experiments initialised is shown in Table 1 below:

Table 1: A description of the different forecast experiments performed to quantify and understand the impact of radiatively interactive prognostic HLO in IFS NWP forecasts.

| Forecast Type | Date Range | Description |
|--|--|--|
| Medium-Range (15 Day) Deterministic | 01/01/2018 – 16/07/2018 17/01/2019 – 23/05/2019 01/07/2019 – 30/06/2020 | Continuously available forecasts at 12-hourly intervals spanning multi-month periods |
| Sub-seasonal (60 Day) Ensemble: SSW Investigation | 05/01/2004 – 05/03/2004 21/01/2006 – 22/03/2006 24/01/2009 – 25/03/2009 09/02/2010 – 10/04/2010 07/01/2013 – 08/03/2013 12/02/2018 – 13/04/2018 | Initialised forecasts on the onset (central warming) date of six PJO-type SSW events |
| Sub-seasonal (60 Day) Ensemble: SSW Radiative Heating Tendency Case Study | 24/01/2009 – 25/03/2009 | Repeated experiment for the January 2009 event with output of radiative heating tendencies |

The impact of the radiatively interactive prognostic HLO scheme is first ascertained by taking the difference between each forecast simulation (interactive minus control) for variables such as temperature, wind (v,w), zonal wind (u) and geopotential height (Z) to diagnose mean changes. Where multiple permutations exist, the forecast ensemble-mean is first calculated, although uncertainty is computed using the statistical power of all ensemble members. Computation of forecast verification metrics such as Root Mean Square Error (RMSE) are however derived here using all ensemble members to ensure statistical robustness, when discerning potential impact of the HLO scheme implementation (made radiatively interactive). As mentioned in Table 1, the sub-seasonal ensemble experiments are all centred on the onset (or central warming) date of six PJO-class SSW events between 2000 and 2020. In the case of the January 2009 SSW event, an additional set of forecast experiments were performed to obtain shortwave (SW) and longwave (LW) radiative heating tendencies from the model, as part of a more detailed investigation of impact mechanisms due to radiatively interactive HLO. A further description of the experiments performed, together with the use of additional datasets (e.g. analysis fields) and forecast skill metrics used is provided in this section.

2.1. Medium-range deterministic forecasts

As displayed in Table 1, a pair of deterministic forecasts out to 15 days ($T+360$ hours), both with and without radiatively interactive prognostic HLO (initialised at repeated 12-hour intervals), were performed over a selection of multi-month intervals between January 2018 and June 2020 (covering all times of the year). For the interactive forecast, the ozone field in each experiment was initialised from CAMS (own-analysis) using an early 47r1 test cycle at Tco399L137 resolution, which corresponds to a horizontal resolution of ~ 32 km (with 137 model levels vertically), and projected according to the HLO scheme specification (detailed in section 1.2). All forecast fields were analysed at 2° horizontal resolution, for a total of 21 vertical levels between 1 and 1000 hPa.

2.2. Sub-seasonal ensemble forecasts

A series of 51-member ensemble experiments, initialised on the SSW onset or central warming dates for six PJO-type SSW events that occurred during northern hemisphere winter between 2000 and 2020 (see Table 1), were performed out to 60 days ($T+1440$ hours). Different to the medium-range pair of deterministic forecast experiments, the ozone field was initialised from ERA-5, albeit at a slightly reduced resolution of Tco319L137 (approximately 36 km in the horizontal). This is seen by the radiative scheme in all interactive forecasts, but only the default zonal-mean monthly-mean climatology field (derived from CAMS), is again visible for all control forecasts. Forecast fields were analysed at select levels ($n = 25$) between 10 and 1000 hPa and gridded at 2° spatial resolution.

2.3. Ancillary datasets

In evaluating the performance of the radiatively interactive HLO scheme, 12-hourly analysis fields from ERA-5 are used for verification of simulated stratospheric ozone, temperature, wind and geopotential height. For calculation of anomalies with respect to climatology, a CAMS-derived zonal-mean monthly-mean climatology is used for ozone, whilst a monthly-mean ERA-5 climatology (between 2000 and 2013) is used for all other remaining variables. For evaluations of the vertical evolution of polar-cap ozone following SSWs, encompassing both the stratosphere and troposphere, analysis data from CAMS was instead used as the representation of ERA-5 ozone below the tropopause is unrealistic (due to parametrization using the CD scheme which is designed only for the stratosphere). Whilst CAMS also made use of the CD parametrization for simulation of stratospheric ozone, prior to Cycle 47r1, a modified version of the Carbon Bond 2005 chemistry scheme (Huijnen et al., 2010) was used in the troposphere, as derived from Chemistry Transport Model 5 (CTM5), that has been integrated into the IFS since 2014 (Flemming et al., 2015). Since operational CAMS Cycle 46r1, prognostic ozone has been used in the radiation scheme, initially parametrized using the CD scheme and later superseded by the HLO scheme as of Cycle 47r1, in the stratosphere.

2.4. Forecast verification metrics

As quantification of the error between forecast and analysis fields (as a best estimate of the true state of the atmosphere), for given parameters such as temperature and wind, a number of forecast skill metrics exist which are typically applied over large spatial areas (e.g. the extratropics or poleward of 20° latitude). The most widely used metric is RMSE, which can be defined as:

$$RMSE = \sqrt{\sum_{j=1}^m w_j (f_j - a_j)^2} \quad (3)$$

For a field of m model grid-points j , f_j is the number of forecast values to be compared with the same number of analysis a_j values, which are then weighted proportionally with the area of the defined grid box (weights w_j sum to 1) (Geer, 2016). A key limitation of this metric when verifying changes in skill between forecasts, however, is the sensitivity to both changes in the mean bias in addition to any change in variability. An alternative metric that is not influenced by the mean is the standard deviation of the error (SDE):

$$SDE = \sigma (f_j - a_j) \quad (4)$$

Where σ is the standard deviation of the difference between all forecast f_j and analysis a_j values over all model grid-points. Another differently defined metric, albeit similar to SDE in that a skill change between two forecasts can be isolated from any change in mean bias, is the anomaly correlation

coefficient (ACC). The World Meteorological Organisation (WMO) defined ACC serves as a measure of forecast ‘success’ with respect to climatology, in terms of pattern correlations:

$$ACC = \frac{[(f - c) - \overline{(f - c)}] [(a - c) - \overline{(a - c)}]}{\sqrt{((f - c) - \overline{(f - c)})^2 ((a - c) - \overline{(a - c)})^2}} \quad (5)$$

Where f is equal to all forecast values, a is equal to all analysis values and c refers to values obtained from climatology (CAM5 for ozone and ERA-5 for temperature and geopotential height). For evaluations conducted here, the latitudinally varying mean bias in forecast with respect to analysis fields is taken into account. For a given spatial area, the ACC quantifies the correspondence (on a scale of -1 to 1) between the forecast and analysis spatial distributions of a given variable. An ACC value of 0.6 is considered to be the approximate threshold for skilful prediction of synoptic-scale features (Krishnamurti et al., 2003; Andersson, 2015), which typically is surpassed between forecast day 8 to 10. Note that ACC is equal to one at the point of forecast initialisation and typically falls away exponentially with increasing forecast lead time until all skill is lost (i.e. ACC = ~0). Therefore, the ACC statistic provides an additional indicator of forecast skill which is useful to compare with both SDE and RMSE.

A fourth and final verification metric used in investigating the impact of the radiatively interactive HLO scheme here is that of the continuous ranked probability score (CRPS); which is applicable only to ensemble (probabilistic) forecasts. It constitutes an extension of the ranked probability score (RPS) to a continuous case (Hersbach, 2000; Matheson and Winkler, 1976; Unger, 1985). The CRPS may be defined as:

$$CRPS = \int_{-\infty}^{\infty} [F(y) - F_o(y)]^2 dy, \quad (6a)$$

Where:

$$F_o(y) = \begin{cases} 0, & y < \text{observed value} \\ 1, & y \geq \text{observed value} \end{cases} \quad (6b)$$

Equation 6b defines the point at which forecast variable y equals the observed value as a cumulative-probability step function that jumps from 0 to 1. Like RMSE and SDE, CRPS is negatively orientated (smaller values represent better skill). Skill is measured in this metric by concentration of probability around the step-function corresponding to the observed value. Thus, the CRPS represents the integrated squared difference between the cumulative density function (CDF) and step-function (Wilks, 2011). Therefore, just like RMSE, comparing the computed CRPS metric for two different forecasts also scores

changes in mean bias, as well as any change in variability. Although this is informative, it is also useful to compare both CRPS and RMSE against both the SDE and ACC skill metrics.

2.5. Radiative heating tendency analysis

To help understand the radiative mechanisms by which the HLO scheme, interactive with radiation, influences temperature (and thus dynamics via the thermal-wind balance relationship), an additional pair of sub-seasonal (60-day) 51-member ensemble forecast experiments (see Table 1) was performed. The 24 January 2009 SSW constituted the most intense and prolonged event on record (Manney et al., 2009a), and therefore presented as an ideal case study for this investigation, in effort of understanding the radiative impact of long-lived PJO-type SSWs. The experiments were again initialised on the central warming (onset) date at Tco319 resolution (~36 km), with 137 vertical levels. Extracted on five model levels between ~10 and 100 hPa, both the SW and LW radiative accumulated heating tendencies were output (with the net tendency computed as the difference) and examined at 1° horizontal resolution. Daily-resolved heating tendencies were subsequently calculated (units: K day⁻¹). All other experimental settings are equivalent to that stated already in section 2.2.

3. Medium-range (0-15 day) results

A selection of results highlighting the performance of the HLO scheme (interactive with radiation) in the IFS are shown in this section, encompassing the medium-range forecast window (up to 15 days). The overall skill change is first assessed over an extended, multi-month period when the solar input strongly differs between hemispheres (section 3.1), before a snapshot case is examined to illustrate the impact of ozone anomalies on temperature and wind during a typical quiescent, summertime state of the stratosphere (section 3.2), when the radiative importance of ozone is maximised. Finally, the short-range impact of radiatively interactive HLO on forecast skill during the major February 2018 SSW is examined (section 3.3), again using the same pair of deterministic forecasts (interactive and control) available at routine 12-hour intervals (details provided in Table 1 and section 2).

3.1. April 2018 to June 2020

Averaged over 11.5 months between April and September (over much of the period from April 2018 to June 2020 during northern hemisphere summer), the zonal-mean change in forecast temperature T and vector wind \mathbf{vw} as a function of pressure (1-1000 hPa), following implementation of the radiatively interactive HLO scheme, is first shown in Figure 3 using the pair of medium-range deterministic forecasts for a lead time of 3 days (T+72 hours) and 7 days (T+168 hours), from a total of 636 to 712 samples. Statistical significance (95% confidence) is inferred using a paired, two-sided t-test, following an applied Šidák correction for 20 independent tests, as a standard method in controlling for multiplicity (e.g. Abdi, 2007). Common to both time steps, is an overall cooling of the mean temperature in the mid-stratosphere (centred around 10-30 hPa), which is concentrated towards the northern (spring/summer) hemisphere. This change is on the order of less than 0.2°C at T+72, increasing to as much as ~0.4°C at T+168 over northern hemisphere mid-latitudes. In contrast, the mean temperature of the lower stratosphere (~70-100 hPa especially) increases by up to 0.3°C by T+168 across all latitudes. This is also found in the upper stratosphere (~1-5 hPa), albeit the mean temperature increase is larger (up to +0.5°C in northern hemisphere mid-latitudes already at T+72, and widely exceeding this everywhere in the northern hemisphere at T+168). As shown by the schematic of temperature biases within the IFS, see Figure 1 in the recently published Polichtchouk et al. (2021) report, the mean stratospheric temperature changes due to radiatively interactive HLO commented on here (statistically significant in each case) translate to a slightly improved representation of the LMS cold bias (~5°C), as well as conversely, a degradation of the mid-stratosphere cold bias (~2°C) and near stratopause warm bias (~10°C). The stratospheric mean changes in vector wind are overall consistent with the differences in mean temperature shown, via the thermal-wind balance relationship.

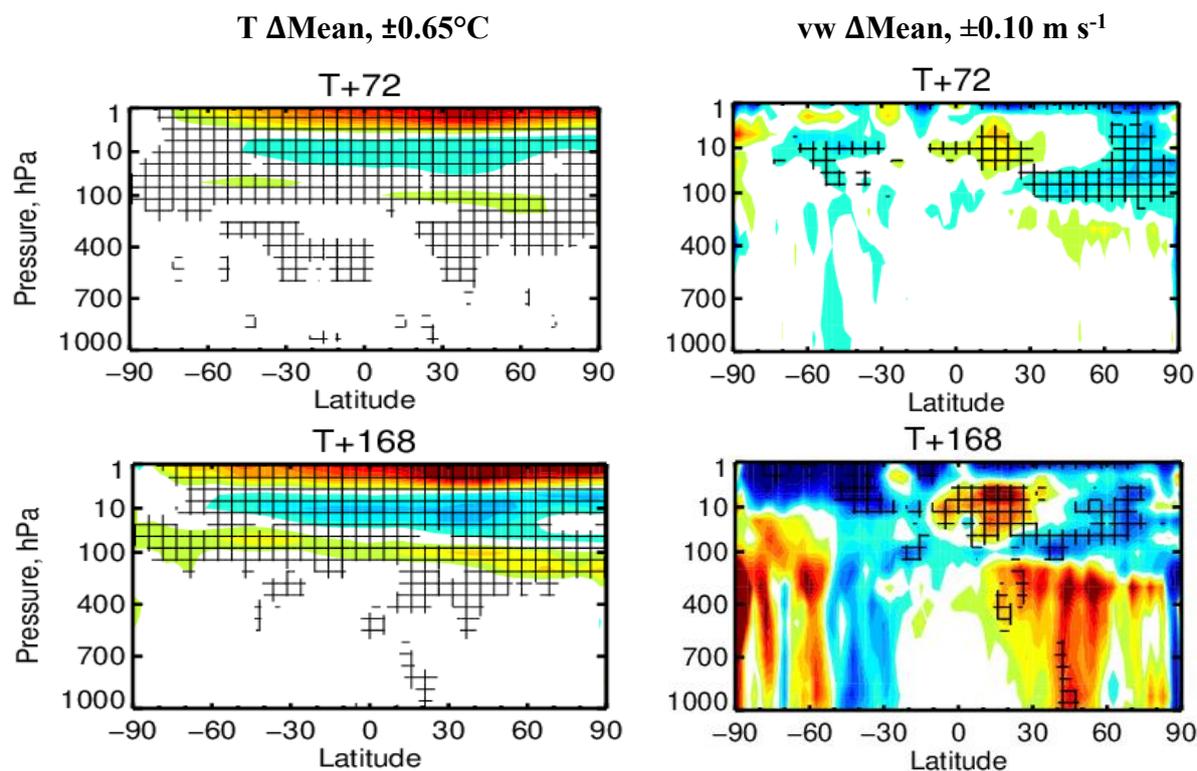


Figure 3: Latitude-pressure distribution of the change in mean forecast error (Δmean) between the interactive and control medium-range deterministic forecasts for (**top row**) T+72 and (**bottom row**) T+168. Results for (**left**) temperature T , and (**right**) wind vw are shown, as verified against analysis data for between 636 to 712 samples. Cross-hatching denotes statistical significance (95% confidence) with a Šidák correction for 20 independent tests.

To further understand how these mean changes translate to forecast skill change, the calculated normalised difference in RMSE at both T+72 (Day 3) and T+168 (Day 7), as a result of implementing radiatively interactive HLO, is next shown for temperature T and vector wind vw (Figure 4). Note that the RMSE forecast verification metric is influenced by both changes to the mean forecast bias, as well as the change in variability. The normalised RMSE difference patterns appear very similar at both time steps shown in each case. An overall increase in temperature RMSE is evident above 10 hPa (of up to 10% by T+168). Over northern hemisphere high-latitudes, an overall reduction (improvement) in RMSE is apparent throughout much of the stratosphere (up to 15%), with a smaller overall improvement over the extratropical southern hemisphere (up to ~5%) between 10 and 100 hPa. For the normalised RMSE change in wind, a relatively large (up to 4%) reduction in RMSE establishes over the extratropical northern hemisphere stratosphere (particularly between 10 and 100 hPa), with a small increase (up to 1%) at low-latitudes (0-30°N) at T+72. A region of slight improvement (~1%) is present over southern hemisphere mid-latitudes during what is largely austral wintertime. At a forecast lead

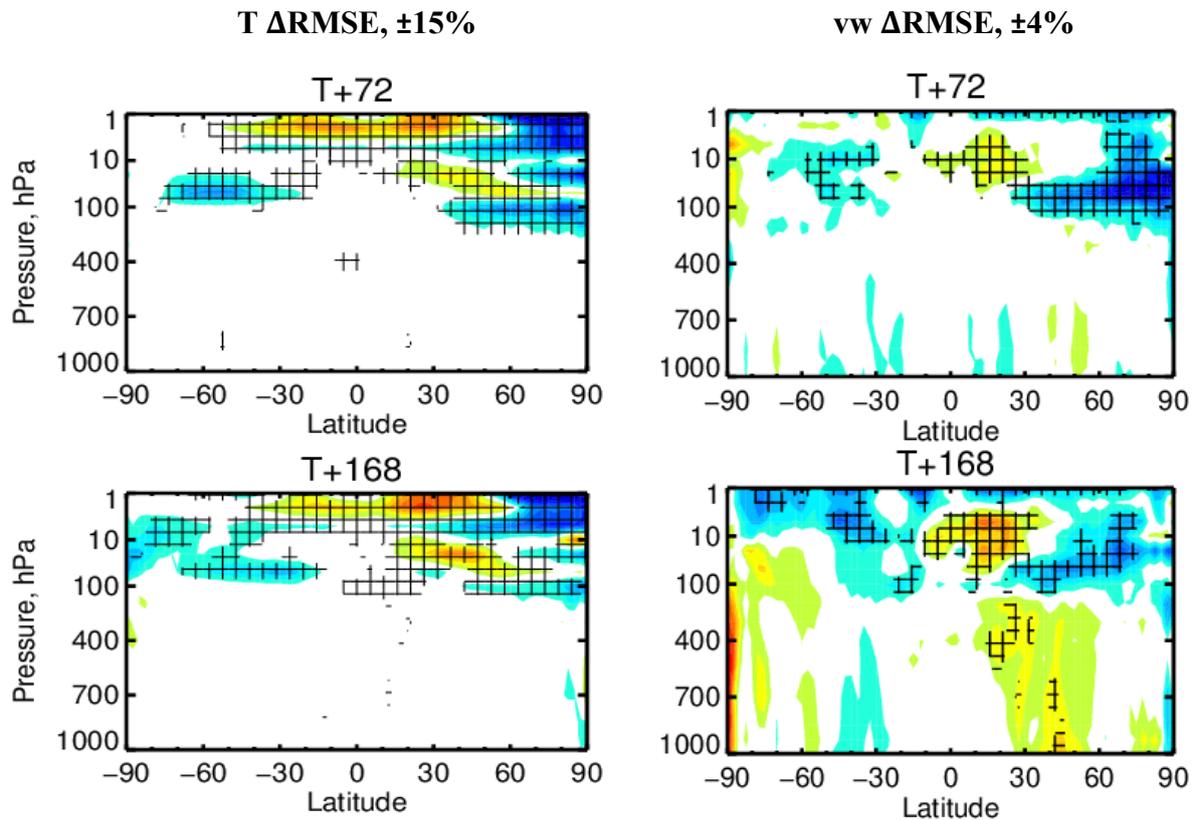


Figure 4: Same as Figure 3 but for the normalised change in RMSE (%).

time of 7 days (T+168), a similar pattern persists in the stratosphere. The RMSE skill improvement is reduced over the northern hemisphere extratropics ($\sim 2\%$), but both enhanced and shifted upward over the southern hemisphere ($\sim 2\%$). The degradation signal at northern hemisphere low-latitudes also increases (up to 2%), which appears to extend into the mid-latitude troposphere (up to 1%), albeit statistical significance is marginal here. In essence, the mean changes shown in Figure 3 explain a significant amount of the improvement/degradation signals in RMSE shown (Figure 4).

In Figure 5, the change in the computed ACC between the deterministic interactive and control forecast is shown for temperature T , as well as for geopotential Z . Note that the ACC difference is not shown for wind as this is a vector quantity. As explained in section 2.4, the ACC metric isolates any skill change from that of changes to the mean bias and is positively orientated (with a perfect score equal to 1, as occurs at the point of forecast initialisation). Over the northern hemisphere extratropical stratosphere ($\sim 10\text{--}100$ hPa), the normalised change in ACC peaks around 10% at T+72, corresponding to overall improvement in the spatial correlation of temperature variability for the forecast with HLO interactive with radiation. A much smaller improvement (up to 2%) is also discernible across much of the southern hemisphere around the same altitude range, with little change in ACC evident elsewhere

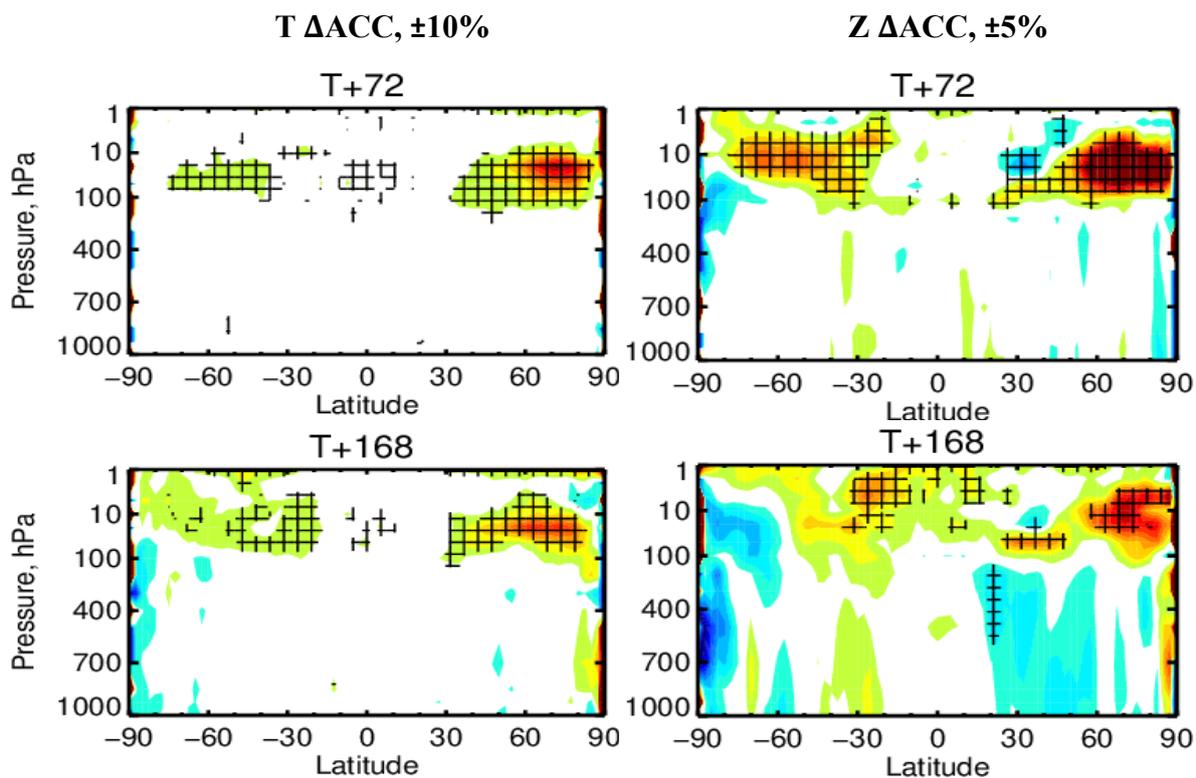


Figure 5: Same as Figure 3 and 4 but for the normalised change in ACC (%) for (left) temperature *T* and (right) geopotential *Z*.

(including all of the troposphere). Both features are again statistically significant, when taking the difference in this score metric, and persist out to T+168 (forecast day 7). The addition of showing the change in ACC for geopotential helps to assess the impact on dynamical skill. As expected, the pattern is similar to that for temperature at either time step, as temperature has a direct influence over this variable in the air column above and thus both these two main features are shifted upwards slightly. The change in ACC over the northern hemisphere for geopotential is up to +5% and nearer +2-3% over the southern hemisphere stratosphere, which reduces slightly by T+168. The lack of statistical significance however inhibits any robust assessment of ACC skill change in the troposphere, which also applies to temperature and wind using any of the skill metrics presented here.

The results discussed in this sub-section are intended to give an overview of the radiatively interactive HLO impact on forecast skill in the medium range, as a function of height and latitude over a sustained period, when the solar input is differs strongly between each hemisphere. The results conform with that expected, in that the impact of HLO, interactive with radiation, is seemingly larger in the northern (summer) hemisphere during this time.

3.2. Relationship between temperature, ozone and wind

The mechanisms by which spatial structures in ozone may influence temperature and in turn wind are next assessed here, using an idealistic example during the quiescent state of the summertime stratosphere over the southern hemisphere (when solar input is high). To understand the impact of anomalies in ozone on temperature (and resultant wind), that may propagate zonally and meridionally around each hemisphere, the spatial fields in ozone and ozone anomaly (with respect to a multi-year

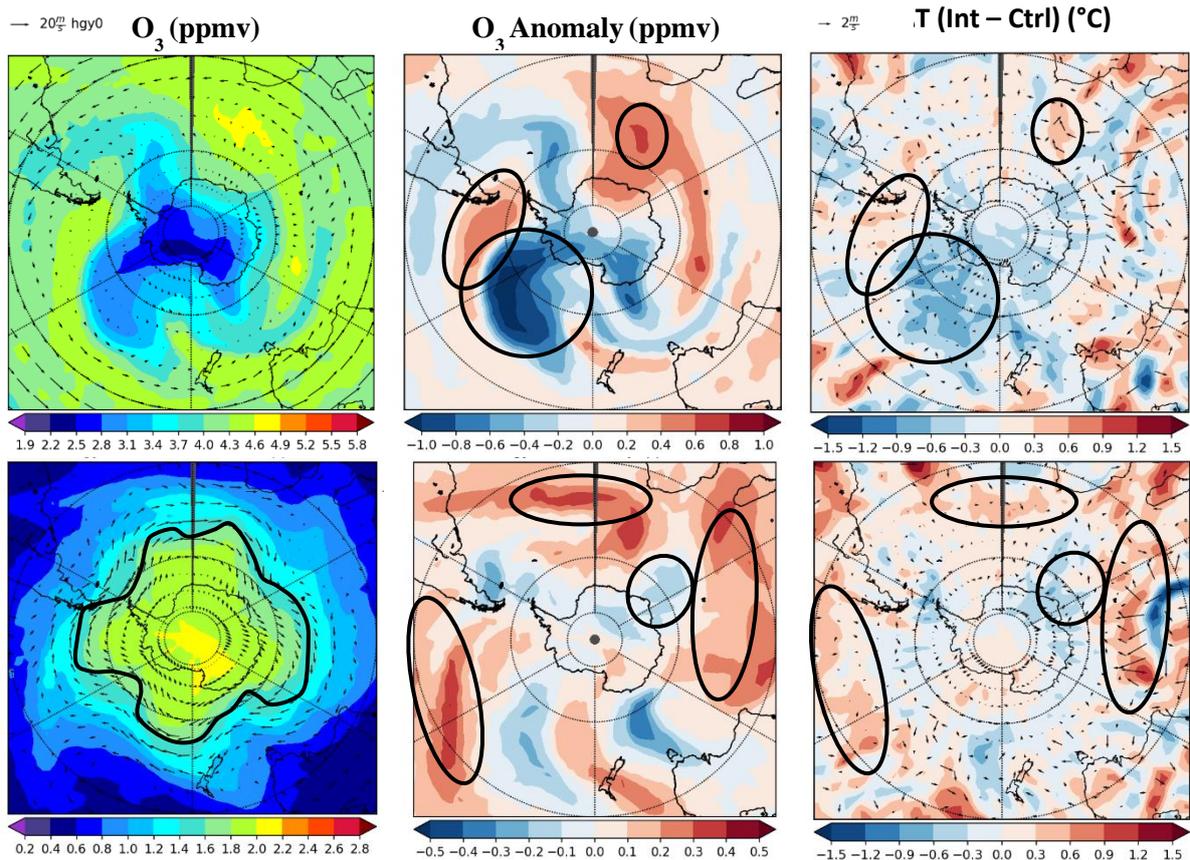


Figure 6: The impact of radiatively interactive prognostic ozone (using the HLO scheme) upon temperature and wind, exemplified during the quiescent state of the summertime southern hemisphere for both (**top row**) 30 hPa and (**bottom row**) 70 hPa. Forecast fields are shown for a lead time of 5 day ($T+120$), initialised on the 2nd of February 2018 at 00 UTC. The forecast distribution of ozone for the interactive experiment is shown (with winds overlaid) (**left**), alongside the corresponding anomaly field with respect to a multi-year zonal-mean CAMS ozone climatology (**centre**). The temperature difference (ΔT) between the interactive and control forecast is subsequently shown (**right**), together with the overlaid wind difference. Regions where same sign anomalies in ozone and ΔT correspond spatially are circled, and a zonal wavenumber-4 pattern in the lower stratosphere (70 hPa) is traced (**bottom left**).

zonal-mean CAMS climatology) parametrized using the HLO scheme are examined for a snapshot case. Alongside this, the difference in temperature (interactive minus control) between the two deterministic, medium-range forecast experiments is shown in Figure 6. Each forecast or forecast difference field is for a lead time of 5 days (T+120), initialised on the 2nd of February 2008 at 00 UTC (i.e. forecast for the 7th of February 2008 at 00 UTC) for 30 and 70 hPa. Although there is large-scale complexity in the temperature difference (ΔT) field, a broad spatial co-location of same sign anomalies in ozone and calculated ΔT can be discerned (as denoted by the circled regions). Although direct radiative feedbacks due to ozone in the interactive experiment therefore likely contributes to the ΔT spatial pattern, it should of course be noted that the ΔT between each forecast will be influenced by indirect feedbacks which are complex and difficult to understand (e.g. perturbations to dynamics operating over a wide range of spatial scales). Additionally, and as expected according to geostrophic balance, the largest changes in wind velocity and vector changes between the interactive and control experiment are strongly associated with regions of largest ΔT . Broadly, there is evidence of wind vector convergence (divergence) in conjunction with positive (negative) values in ΔT that would be expected when calculating differences in temperature and winds between these two forecast simulations.

3.3. February 2018 SSW: Polar-cap (60-90°N) evolution

Skill changes due to radiatively interactive HLO in the medium range are next assessed over the northern hemisphere polar-cap (60-90°N) region around this time, corresponding with the build-up, onset and aftermath phases of the February 2018 SSW event, using the pair of deterministic forecasts

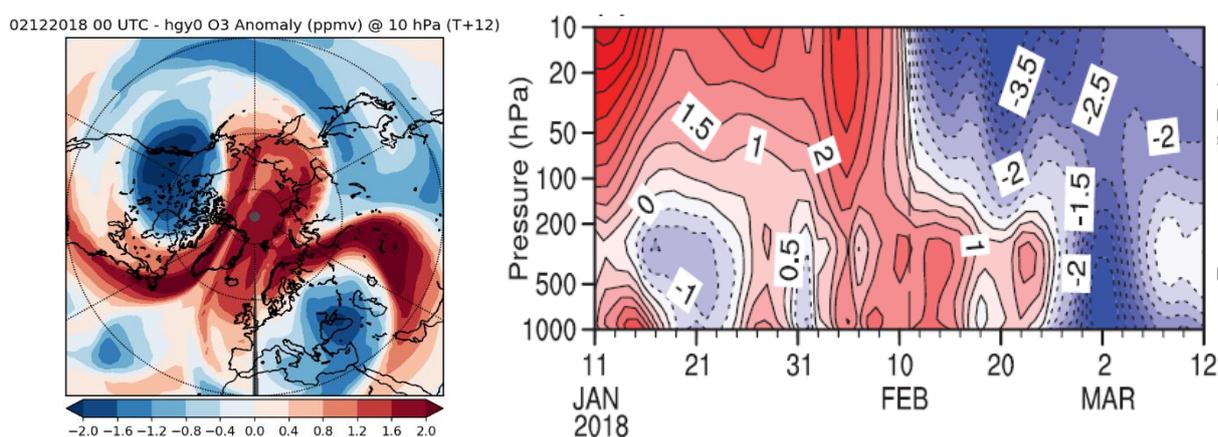


Figure 7: The ozone anomaly distribution (with respect to the CAMS multi-annual climatology) at 10 hPa (**left**), as reconstructed from the ERA-5 analysis on the SSW onset (central warming) date of 12th February 2018 at 12 UTC. The evolution of the northern annular mode (NAM), as taken from Rao *et al.* (2019; Figure 1c), spanning from a month before to a month after this date is additionally shown (**right**) to highlight the abrupt shift in the dynamical circulation regime as a result of the SSW.

examined so far. On the 12th of February 2018, a major SSW occurred in the northern hemisphere. The SSW was classified as a PJO event, as well as a dynamical SPV split event. The dynamical classification is illustrated in Figure 7 (left) from the ERA-5 ozone anomaly analysis field at 10 hPa on this day. Poleward advection of anomalously high ozone from mid-latitudes into the polar-cap (60-90°N) region is indicated on the forward flanks of the two main daughter vortices (captured by the regions of negative ozone anomalies over western Canada and south-eastern Europe respectively). In terms of the dynamical impact of the event, this is illustrated by the height-resolved northern annular mode (NAM) index evolution (Figure 7; right) before and after the event. The impact on the troposphere can be seen to manifest approximately 10 days after the SSW onset (~22nd of February), characterised by negative values exceeding -2 units of the NAM index.

The evolution in the forecast difference in polar-cap averaged temperature, between the 15-day interactive experiment and control forecast, was first explored for four key dates associated with different stages of the event, in conjunction with the ozone anomaly and forecast minus analysis evolution in temperature (not shown). The forecast initialisation dates (at 12 UTC) included: 2nd February (10 days before the event onset); 12th February (event onset date); 22nd February (10 days after the event onset) and 4th March (20 days after the event onset). The results indicated that radiatively interactive HLO reduced the mean forecast error in stratospheric temperature leading up to the event, by as much as 30-40%, but with largely neutral impact for forecasts initialised at the later times examined. For the forecast initialised 20 days after the event onset however, a local improvement in temperature representation was apparent in the lower stratosphere (~70-100 hPa) due to radiatively interactive HLO, in conjunction with the residence of a positive lower stratospheric anomaly in ozone. The inference is that such feature provides a direct local radiative feedback, when shortwave heating becomes increasingly important with the transition into spring. The deterministic nature of the comparison however precludes a robust assessment of these findings and is thus not presented here. The performance of the radiatively interactive HLO scheme in the context of northern hemisphere PJO-type SSW events is in any case investigated in section 4, for a suite of sub-seasonal ensemble forecasts initialised on the onset dates for six different events between 2000 and 2020. Nonetheless, the repeat availability of medium-range (15-day) deterministic forecasts helps to provide an indication of how the HLO scheme performance varies during the life cycle of an SSW.

To understand whether radiatively interactive HLO lead to a better overall representation in temperature as the February 2018 SSW event evolved, the area-weighted RMSE evolution for forecast day 5 (T+120 hours), computed over the polar-cap (60-90°N) region, is displayed in Figure 8 for each experiment, together with the normalised difference (interactive minus control). The period spans the 30 days leading up to the event (13th of January) through to 90 days following the SSW onset (10th of May). Note that to reduce noise, a 7-day moving average is constructed from the initial calculated RMSE field

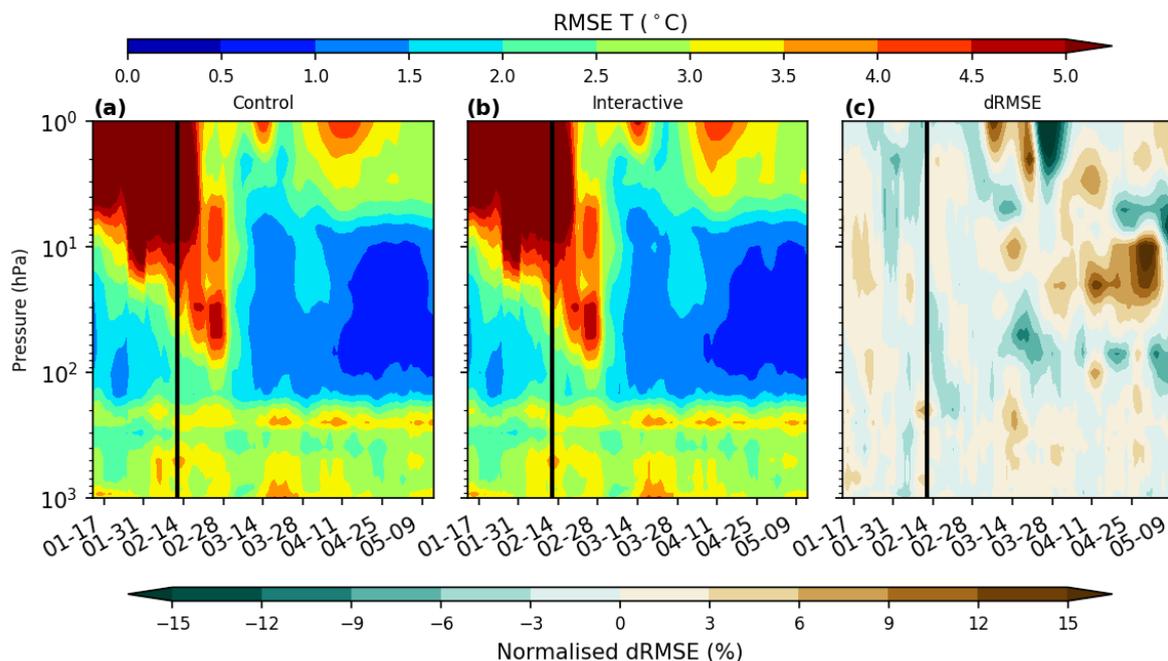


Figure 8: Polar-cap ($60\text{-}90^\circ\text{N}$) area-weighted average 5-day ($T+120$) RMSE evolution for (a) the control experiment, (b) the interactive experiment and (c) the interactive minus control forecast difference from the 13th of January 2018 (30 days prior to the SSW central warming date as denoted by the black vertical line) to the 10th of May 2018 (90 days following the SSW onset). These dates correspond to that of the verifying analysis (constructed from the interactive experiment at $T+0$ lead time). Note that to reduce temporal noise, the RMSE evolution shown here is constructed from a 7-day moving average.

which is displayed here. The evolution in the RMSE at this lead time is shown for both experiments; largest earlier in the period ($>5^\circ\text{C}$) as the SSW onset invokes a large change in stratospheric temperature, which gradually propagates downwards. After around 15 days beyond the central warming date (end of February), RMSE falls sharply throughout much of the stratosphere as the SPV only slowly reforms at higher altitudes and the lower stratosphere remains strongly perturbed with respect to climatology ($>2^\circ\text{C}$). This reflects the dearth of dynamical heating as planetary wave propagation is inhibited by the SSW-induced wind reversal, with stratospheric temperature controlled by slow radiative relaxation timescales (Hitchcock et al., 2013; de la Cámara et al., 2018b). The RMSE evolution in the troposphere on the other hand remains highly variable over this period, with no significant changes apparent as the SSW evolves. The difference in RMSE (dRMSE) between the interactive and control forecasts shows an overall slight improvement. RMSE is reduced by between ~ 3 and 6% in the upper- to mid-stratosphere (1-10 hPa) prior to the onset date (12th of February) and in the 10 days after this date in the lower stratosphere ($\sim 30\text{-}200$ hPa). Approaching ~ 3 weeks after the event onset (i.e. early March), larger changes in RMSE are evident, but this is strongly altitude and time dependent. In the

lower stratosphere (between ~70 and 100 hPa), a coherent signal for a reduction in RMSE by up to 10% due to radiatively interactive HLO is however evident. This signal would correspond with the residence of a long-lived positive anomaly in ozone, as identified from earlier forecast evaluation (not shown), and it could thus be the case that inclusion of radiatively interactive HLO is serving to improve mean temperature representation within the IFS in this region.

The deterioration in other regions can be attributed to an increase in the mean temperature bias, since the RMSE metric is sensitive to this aspect as well as any variability associated skill change, as subsequently shown in Figure 9 for the corresponding temperature ACC evolution. Despite this positive inference regarding the performance of the HLO scheme, the variability aspect of the skill change is less clear for the ~70-100 hPa region. Although, the pattern correlation in temperature is improved leading up to the event onset in the mid-stratosphere (equating to around 2%), the signal in the lower stratosphere beyond 20 days after the onset is largely neutral. The signal is inconclusive ($>\pm 2\%$) in the troposphere throughout, as well as in the mid-stratosphere after 30 days following the onset, which may reflect interference from indirect dynamical feedbacks. Any direct variability aspect of skill change due to HLO, interactive with radiation, can easily be swamped by such feedbacks, which become readily significant with increasing forecast lead time.

To ascertain representation of ozone, simulated using the HLO scheme, during other similar events (i.e. PJO-type SSWs), similar and more detailed evaluations are next shown in section 4 for a series of sub-seasonal ensemble forecast experiments, initialised on the central warming dates of six different events that occurred between 2000 and 2020 (including that of the February 2018 event once more).

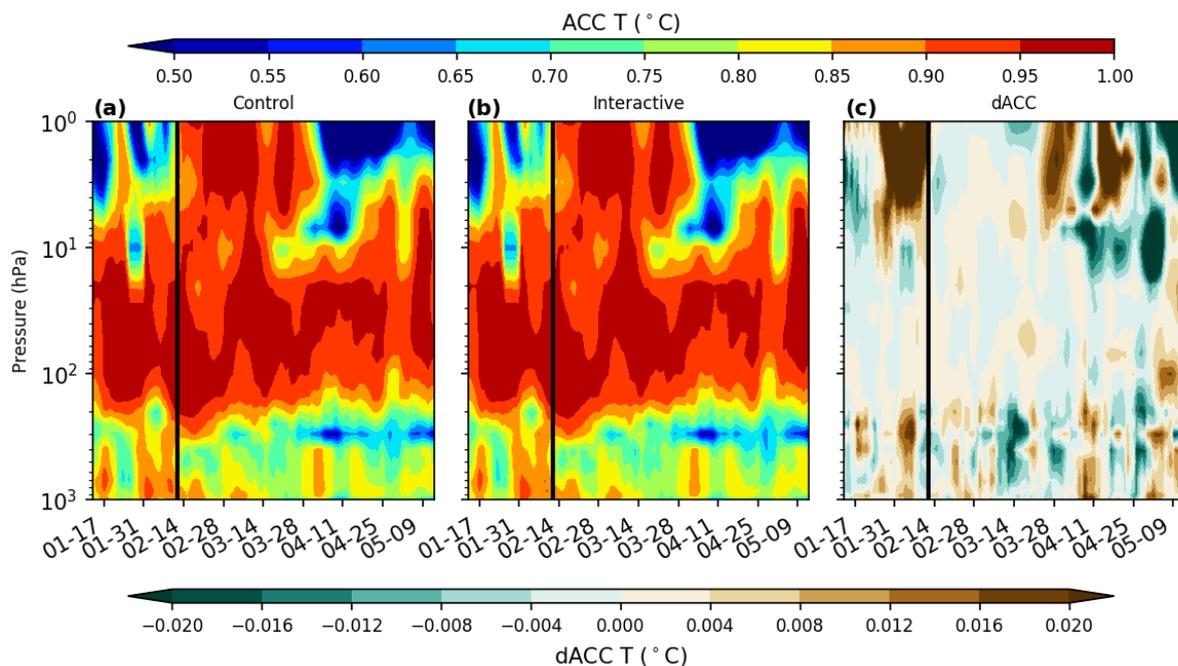


Figure 9: Same as Figure 8 but for the ACC skill metric.

Investigation of the radiatively interactive HLO scheme performance during the quiescent summertime state of the southern hemisphere stratosphere at the same time is also provided when the solar input is larger (and dynamical activity is weak) for comparison.

4. Sub-seasonal (0-60 day) results

Findings of the performance of the radiatively interactive HLO scheme on sub-seasonal timescales are presented in this section, using a probabilistic (ensemble) forecasting approach, to compare and contrast with the medium-range findings in section 3. Of particular interest over this forecast time range, is the propensity for HLO, interactive with radiation, to improve the stratosphere-troposphere coupling representation following major northern hemisphere midwinter SSWs. Since the inclusion of radiatively interactive ozone has been linked to the persistence of surface anomalies after such events (e.g. Haase and Matthes, 2019; Oehrlein et al., 2020), the tropospheric skill gain is likely to manifest at timescales longer than the medium range. Therefore, a series of forecast experiments have been performed out to 60 days, to investigate skill change in response to radiatively HLO; centred on the central warming (onset) dates of six individual PJO-type SSW events between 2000 and 2020 (see Table 1 and section 2.2 for details of the experimental setup). The ensemble approach additionally helps to address the issue of sampling variability that limited the robustness of the results in section 3. The composited results of the radiatively interactive HLO impact on polar-cap temperature, over all events, are first shown in section 4.1, together with the January 2009 and January 2013 individual events subsequently in section 4.2 and 4.3 (to highlight variance in the signals between events). Changes in both the temperature ACC and CRPS skill metrics due to radiatively interactive HLO are then examined here globally in section 4.4, during the forecast intervals sampled, to ascertain the impact as a function of latitude (solar input). This is again provided as composite over all events, and for the individual events January 2009 and January 2013.

4.1. Composite Approach

The polar-cap (60-90°N) averaged evolution of the interactive forecast temperature anomaly (computed with respect to a monthly-mean ERA-5 climatology over the period 2000-2013 inclusive), together with the difference in interactive temperature (with respect to both ERA-5 analysis and control forecast), is shown in Figure 10 as a function of pressure and time, composited for the six different PJO-type SSW events (between 2000 and 2020) in which each set of forecasts (51 ensemble members) were initialised (on the central warming date). The composited temperature anomaly evolution (panel a) captures the large and abrupt warming in the stratosphere ($>10^{\circ}\text{C}$ between 10 and 30 hPa at the start of the forecast), which gradually erodes in the first month (slowest in the lower stratosphere due to longer radiative relaxation timescales), to be replaced by anomalously cold temperatures thereafter in the mid-stratosphere ($>6^{\circ}\text{C}$ from 10-30 hPa) that descends slowly out to 60 days. Within days, the establishment of the characteristic polar LMS cold bias in IFS occurs (see Polichtchouk et al., 2021), centred around the 200 hPa level (panel b). However, this does not maximise until 25 days onwards, when the

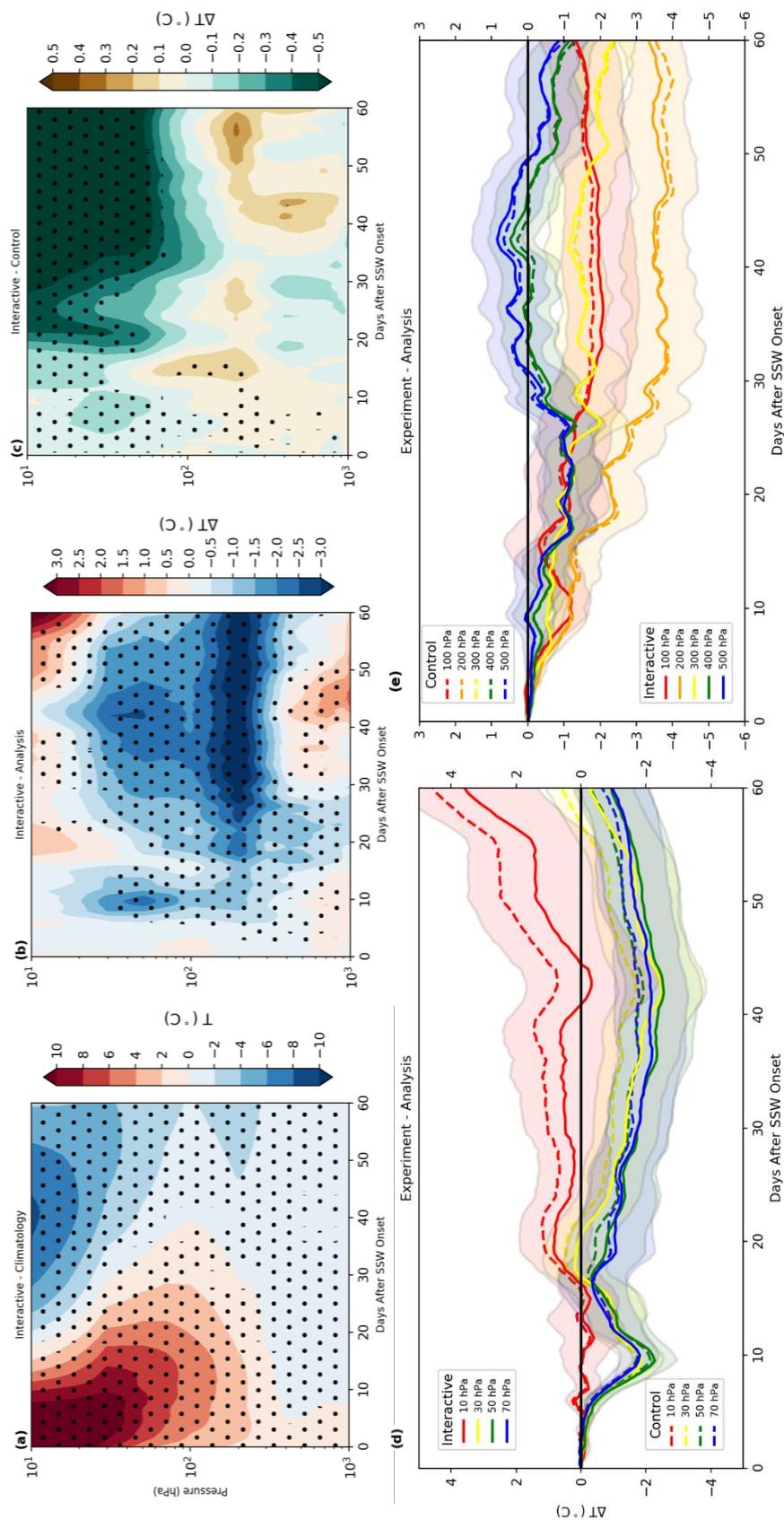


Figure 10: Evolution of the ensemble-mean, polar-cap ($60\text{-}90^{\circ}\text{N}$) area-weighted average of (a) the interactive experiment temperature anomaly with respect to ERA-5 climatology; (b) the temperature difference between the interactive experiment and ERA-5 analysis; (c) the interactive minus control experiment temperature difference; (d) the same information as (b) but for both experiments (for levels 10, 30, 50 and 70 hPa); and (e) the same as (c) (for levels 100, 200, 300, 400 and 500 hPa), following the composited SSW central warming (onset) date over all six events. Stippling in (a) and (b) denotes statistical significance according to a one-sided student's t -test and a two-sided t -test in (c). Solid (dashed) lines in (d) and (e) represent the event/ensemble-mean interactive (control) experiment difference with respect to ERA-5 analysis, with the 1σ of the ensemble spread for the interactive forecast additionally shaded.

difference approaches $\sim 3^{\circ}\text{C}$. A general cold bias indeed establishes with time across much of the lower stratosphere (between 30 and 200 hPa), of as much as 2°C after 30 days lead time. In the mid-stratosphere (close to 10 hPa), the agreement in interactive simulated temperature with ERA-5 is close, but is slightly overestimated if anything after around 20 days into the forecast (up to 1°C). The temperature in the troposphere on the other hand is slightly underestimated between 10 and 20 days (up to 1°C) and overestimated by a similar magnitude between 40 and 50 days. These features mentioned are all robust, according to the two-sided t-test performed (as measure of statistical significance). Line graphs of these differences are also shown for both stratospheric (10-70 hPa) and UTLS (100-500 hPa) levels (panels d and e), together with the computed 1σ range as an indicator of ensemble spread, which largely corroborates this assertion. The radiative impact of HLO can subsequently be seen to slightly improve the temperature representation in the region most affected by the IFS cold bias (~ 200 hPa), warming this region by up to ~ 0.2 - 0.3°C . However, a statistically significant cooling of greater than 0.5°C between 10 and 70 hPa instead worsens the cold bias above this region, whilst improving locally the projected temperature evolution near 10 hPa. The tropospheric signal from using HLO, interactive with radiation, is also one of slight deterioration (by up to $\sim 0.2^{\circ}\text{C}$). It should be noted the change in

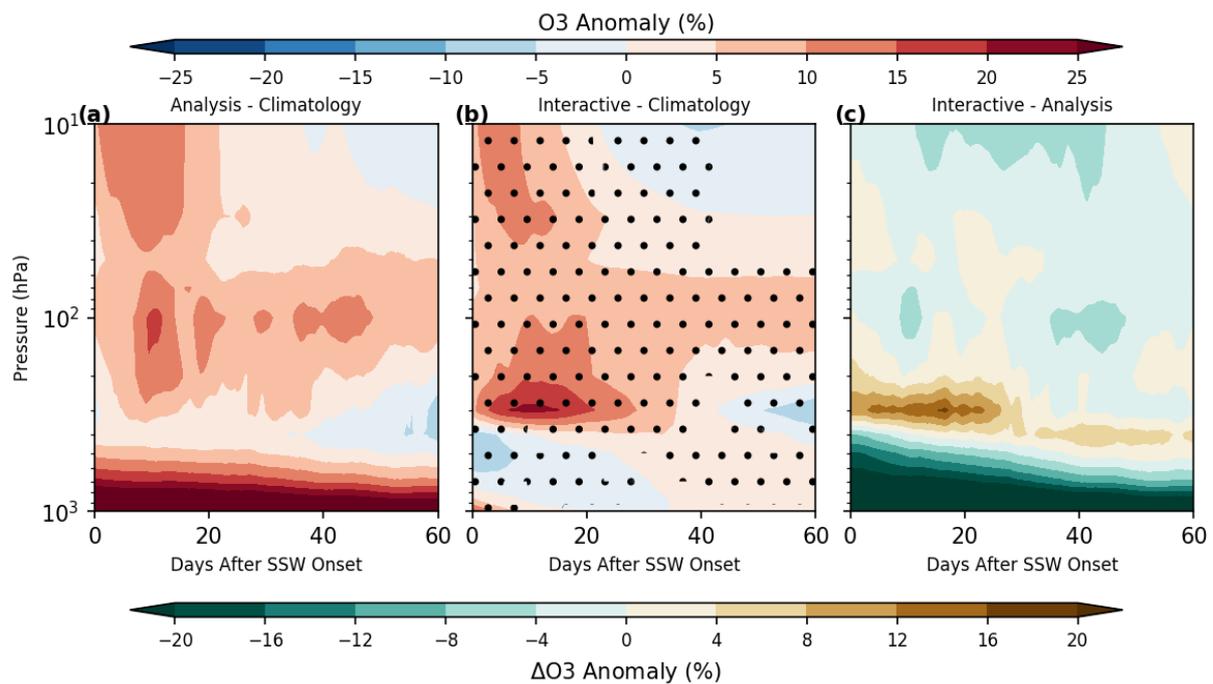


Figure 11: Evolution of the ensemble-mean, polar-cap (60-90°N) area-weighted average of (a) the CAMS reanalysis (2004-2018) ozone anomaly (with respect to the multi-annual CAMS climatology); (b) interactive ozone minus the same climatology and; (c) the interactive ozone normalised difference (%) with respect to the CAMS analysis, following the composited SSW central warming (onset) date over all six events. Stippling in (b) denotes statistical significance according to a one-sided students t-test.

temperature due to radiatively interactive HLO is significantly smaller than the forecast difference with respect to the analysis for each experiment, by a factor of ~5-10.

The CAMS reanalysis and interactive forecast ozone normalised anomaly (%) evolution is shown in Figure 11, together with the difference, again as a function of pressure and forecast lead time for the polar-cap average over all events and ensemble members. The agreement between the forecast and reanalysis is remarkable during the first 30 days, with both a positive anomaly in ozone centred between 10 and 50 hPa (~5-15%) and between 100 and 300 hPa (up to 25%). However, the persistence of both features is underestimated by the interactive forecast by several days, signalling a quicker recovery to nearer climatological values in ozone overall following these events using the radiatively interactive HLO scheme. Beyond 30 days into the forecast, a residual positive anomaly in ozone of up to 10% is maintained between ~70 and 150 hPa according to both the interactive forecast and CAMS. The difference field (panel c) highlights the overall slight underestimation of the ozone anomaly both earlier and later in the forecast, except near the tropopause where the HLO scheme overestimates the anomaly in ozone in the first 30 days by ~10-15%. Note that the representation of ozone in the troposphere from the CAMS reanalysis is much more realistic than that derived from the CAMS multi-year climatology, so the evolution of the ozone anomaly (panel a) can be disregarded as meaningful. The tropospheric evolution in the HLO minus CAMS reanalysis evolution (panel c) is similarly an artefact of the ozone scheme design. Linear ozone schemes such as HLO are not suited for use in the troposphere, with a typical underestimation by a factor of ten over long integrations (Hogan et al., 2017).

Because each SSW event has invariably a different character, the same evaluations shown here are next replicated for a couple of the individual events included in this composite analysis, to highlight variance in the signals between events. In section 4.2, the evolution in polar-cap (60-90°N) temperature and ozone is shown for the 24th of January 2009 event, followed by section 4.3 which encompasses the 6th of January 2013 event. Both events were notably pronounced, with significant impact upon the dynamics of the lower stratosphere and indeed the troposphere noted, and thus these events provide ideal case studies to next examine.

4.2. 24 January 2009 event

The evolution in Arctic temperature is shown in Figure 12, for the interactive experiment with respect to both ERA-5 climatology and analysis (panels a-b), and relative to the control experiment (panel c), out to 60 days lead time following the 2009 SSW central warming date (24th of January). The evolution in the mean forecast temperature anomaly and forecast error closely resembles that for the composite approach (Figure 10), with a few subtle differences. Additionally, the temperature difference (interactive minus control experiment) evolution is overall quite similar, with a general warming (cooling) impact due to radiatively interactive HLO in the lower (middle) stratosphere. The difference

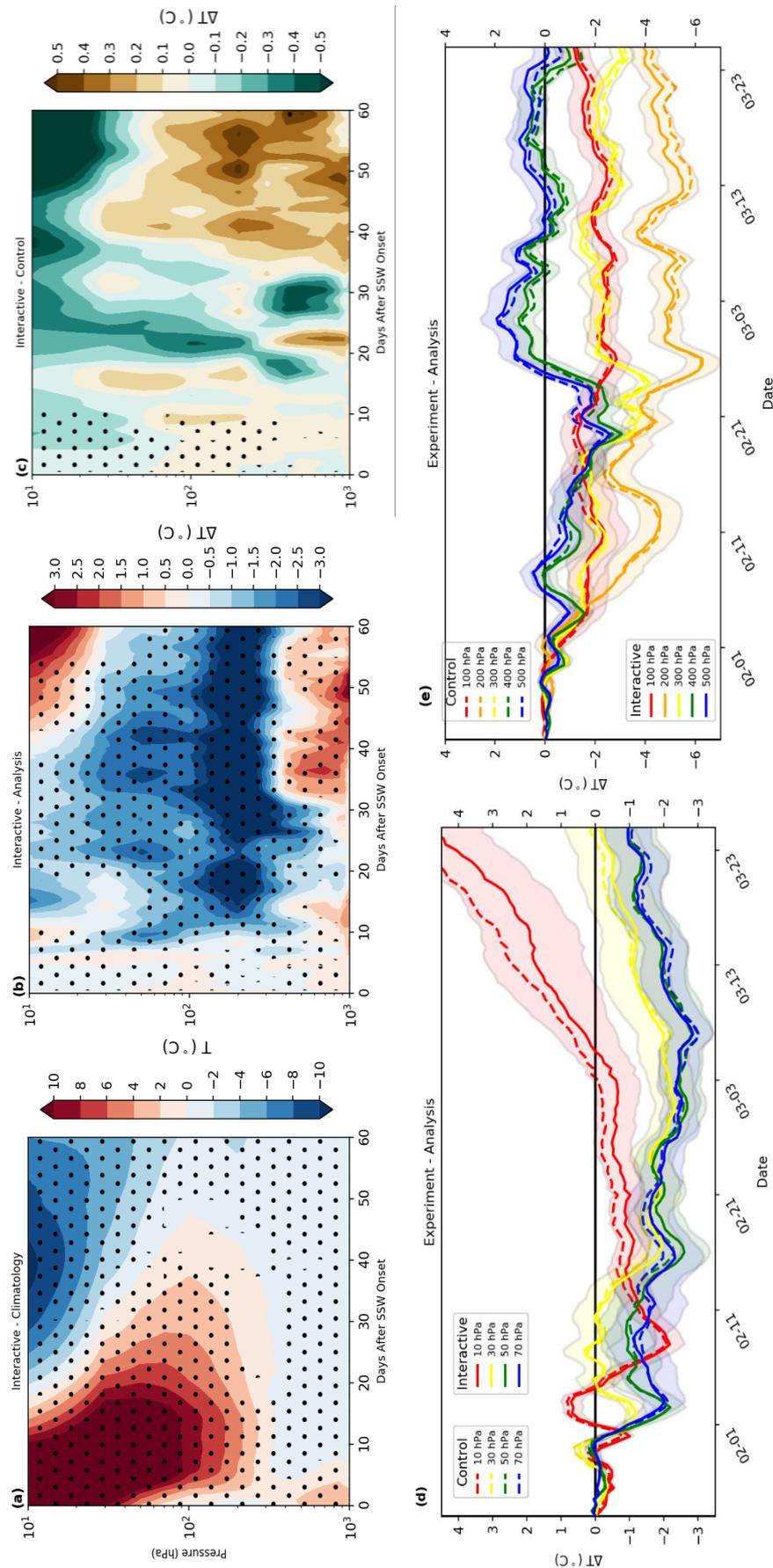


Figure 12: Same as Figure 10 but for the January 2009 SSW event only.

is very small in the first two weeks of the forecast (slight reduction in the lower stratospheric cold bias due to the HLO scheme by $\sim 0.1\text{-}0.2^\circ\text{C}$), with a short period of cooling ($\sim 0.3^\circ\text{C}$) for the interactive case afterwards. This feature is one of degradation as the cold bias throughout the stratosphere at this time is enhanced. After 30 days into the forecast, the story is again one of overall improvement in the lower stratosphere as the HLO scheme radiatively warms this region by up $\sim 0.5^\circ\text{C}$ (compensating slightly for the cold bias). However, it is worth remembering that the bias still exceeds -3°C between ~ 150 and 300 hPa, even in the interactive case. The expanse of the 1σ interval for select pressure levels in the stratosphere (panel d) and UTLS (panel e) region, as representation of the ensemble spread for the interactive forecast, indicates that the sign of differences is generally robust throughout the forecast. The warming due to radiatively interactive HLO also extends into the troposphere, but this translates to skill degradation as the mean forecast bias is enhanced. In the mid-stratosphere (close to 10 hPa), a radiative cooling effect due to the HLO scheme persists from 20-60 days into the forecast which serves to enhance a cold bias initially before acting to counteract a warm bias after 40 days (skill improvement).

The evolution in the interactive ozone anomaly following the 2009 SSW event, verified against the CAMS reanalysis, again shows good agreement throughout the stratosphere (Figure 13). The two main regions characterised by a positive ozone anomaly during the first 30 days of the forecast ($10\text{-}70$ hPa and $100\text{-}300$ hPa) is consistent with that seen in the composite approach (Figure 11), albeit the anomaly tends to be larger: $15\text{-}20\%$ between 10 and 70 hPa and widely exceeding 25% between 100 and 300 hPa. Between 30 and 60 days, the anomaly becomes slightly negative towards 10 hPa (up to 10%), in

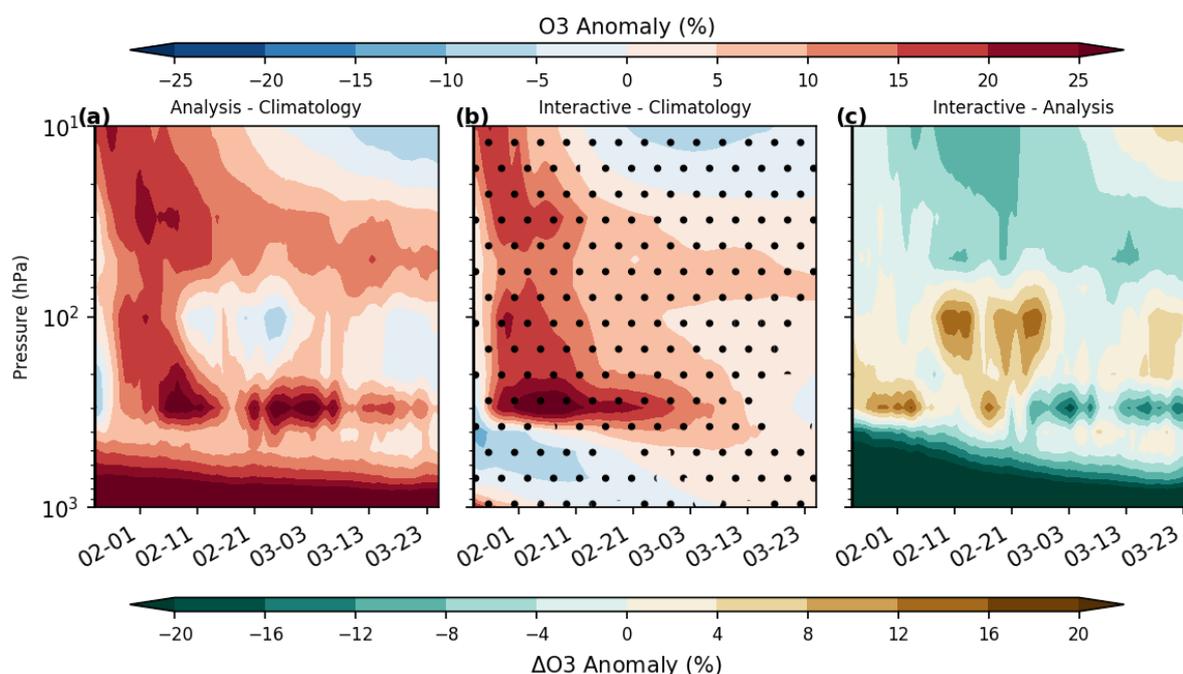


Figure 13: Same as Figure 11 but for the January 2009 SSW event only.

association with SPV recovery, with the core of the positive anomaly sinking down towards 50-70 hPa (still as high as 10-15%). The ozone anomaly in the LMS (100-300 hPa) erodes gradually but remains positive out to 60 days. Above 70 hPa, the HLO scheme still appears to underestimate the abundance in ozone (panel c), but typically overestimates in the lower stratosphere (especially during the first 30 days), except near to the tropopause thereafter. In fact, much of the stratospheric ozone anomaly structure is underestimated by up to 10%, which was also a feature when compositing for all events, as was too the more localised overestimation in the LMS during the first 30 days.

4.3. 07 January 2013 event

To compare such signals for both the composite approach (section 4.1) and the 2009 SSW (section 4.2) with another event, verification of the polar-cap average (60-90°N) temperature and ozone forecast error is furthermore examined following the January 2013 event (out to 60 days lead time). The interactive temperature versus ERA-5 climatology and analysis (panels a-b), as well as the difference with respect to the ensemble-mean of the control experiment (panel c), evolution is displayed in Figure 14. The only major difference in the temperature forecast error evolution (panel b), compared with either the 2009 event or the composite approach for all events, is the development of an overestimation in temperature (between 10 and 25 hPa) a week or so into the forecasts, which then persists out to 60 days (up to 3°C). The development of a cold bias in the lower stratosphere is very consistent with earlier findings (largest near 200 hPa where the interactive experiment is over 3°C colder). The interactive forecast ensemble spread (1σ) again indicates that the sign of the difference with respect to ERA-5 analysis is often robust throughout the forecast, for both stratosphere (panel d) and UTLS (panel e) levels. The tropospheric evolution is also very similar with a cold (warm) bias between 10 and 30 days (30 and 60 days). Although the exact reason for this is unclear, the timing of the former would be consistent with downward coupling influence from the stratosphere (propagation of cold bias signal into the troposphere) and likely cessation thereafter (tropospheric model feedback).

An overall consistent evolution of the difference in mean temperature between the two forecasts is once more evident for the January 2013 event, with a warming (cooling) or the lower (middle) stratosphere due to radiatively interactive HLO, translating to a local reduction (enhancement) of the inherent cold bias within the IFS over the polar-cap during winter. The cold bias in the LMS is generally slightly reduced (~100-300 hPa) after around 10 days into the forecast, by as much as 0.3 to 0.5°C, but enhanced above this region up to around 30 hPa (particularly after 30 days when this is larger than 0.5°C). A warm bias observed near to 10 hPa for this event is however again reduced by a similar magnitude. The radiative impact of HLO in the troposphere is rather mixed but is generally reflective of slight deterioration (improvement) between 20 to 40 days (40 to 60 days). It is likely that much of the changes discussed in this section are associated with the placement of anomalies in ozone following the event.

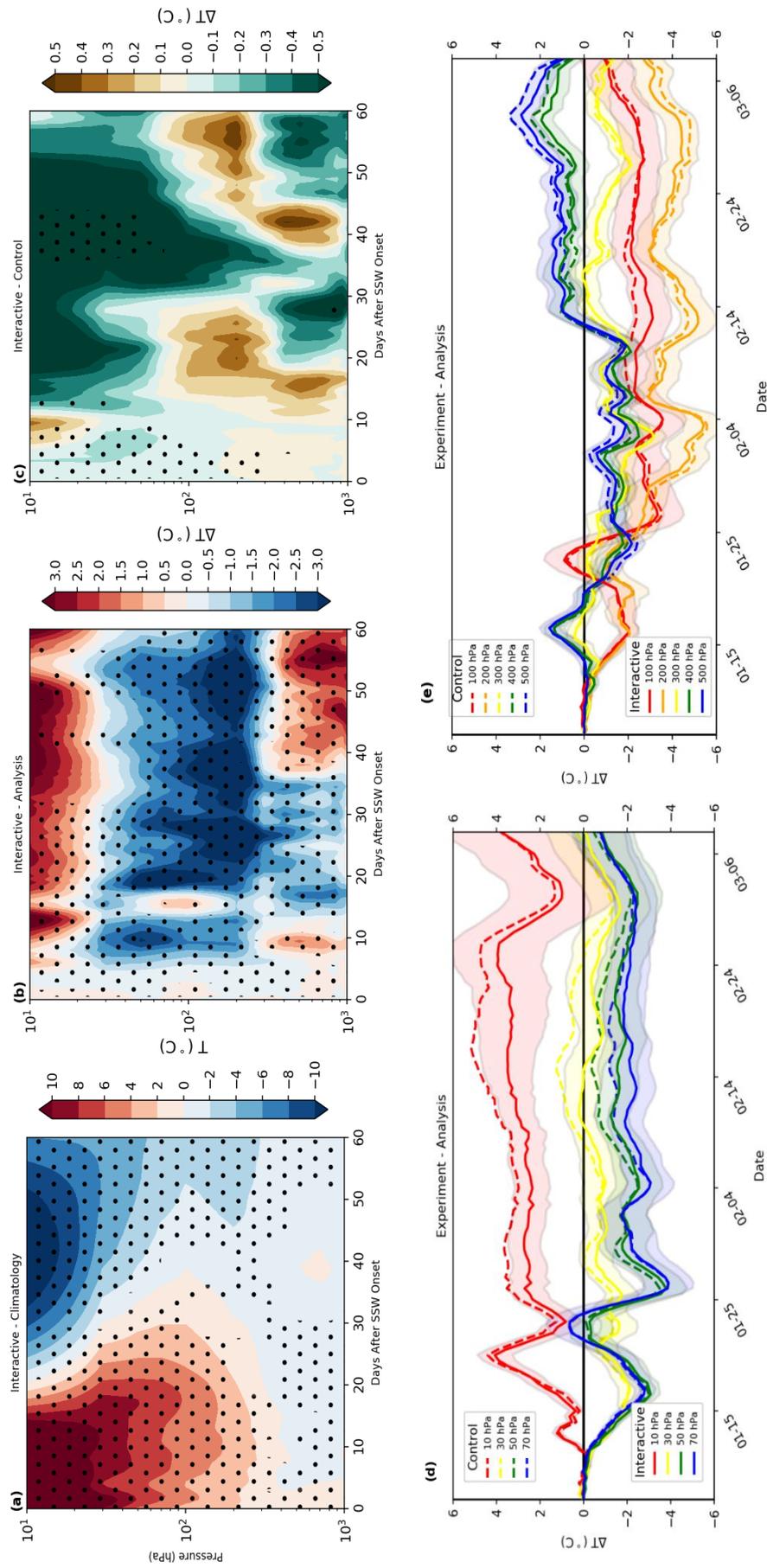


Figure 14: Same as Figure 10 and 12 but for the January 2013 SSW.

To understand this in full however, more detailed investigation of the interactions between ozone, radiation and temperature is necessary and will indeed be the focus of later section 5 for the January 2009 SSW case study.

The ozone anomaly evolution for the January 2013 SSW again shows two distinct regions of positive anomalies in ozone following the event (Figure 15). The feature between 10 and 50 hPa is however a little weaker than evident following the 2009 SSW (Figure 13) and the composite for all events (Figure 11), which also decays fully by 30 days lead time, according to both the CAMS reanalysis and the ensemble-mean of the interactive forecast. Similar to 2009 and for all events, a much more pronounced relative anomaly in ozone establishes in the LMS (~100-300 hPa) for the entirety of the forecast, which is strongest in the first 30 days (>25%) but substantially weaker thereafter (~5-10%). Despite the overall good performance of the HLO scheme in terms of the mean polar-cap ozone evolution, CAMS implies that the real anomaly in ozone is often some 5 to 10% greater throughout much of the forecast. However, an overestimation by the HLO scheme is again signalled near the tropopause (~200-300 hPa) during the first month of the forecast (by up to ~15-20%) as evidenced in panel c.

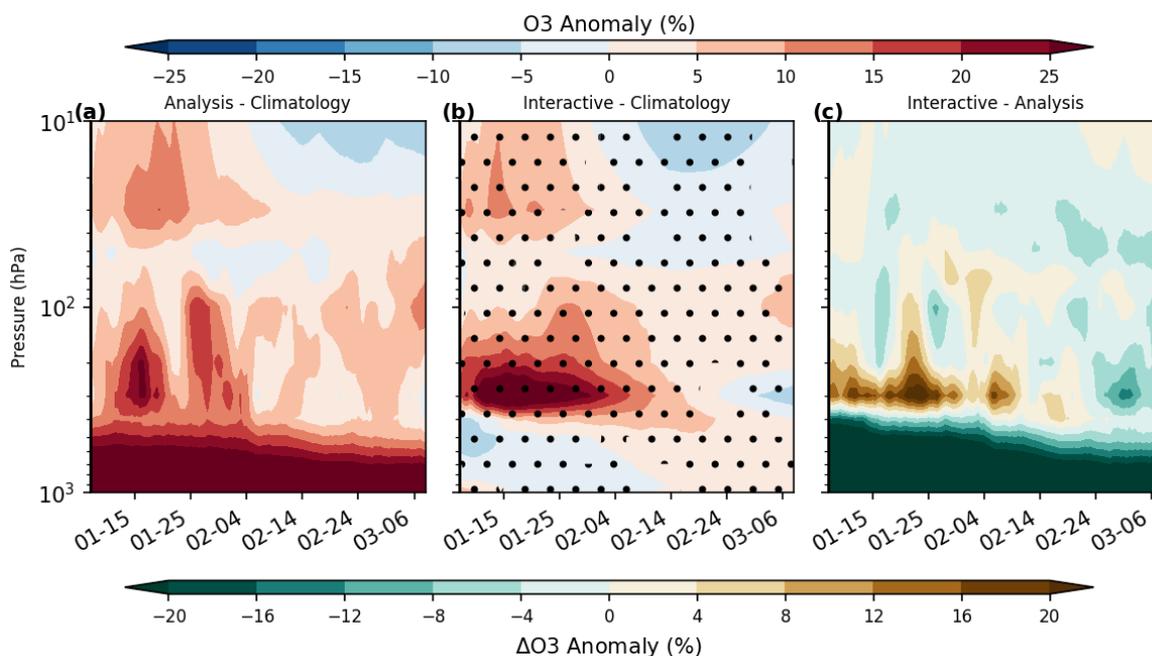


Figure 15: Same as Figure 11 and 13 but for the January 2013 SSW.

4.4. Forecast verification

To further evaluate the performance of the HLO scheme for the composite approach over all six PJO-type SSW events (2000-2020), in addition to the two individual events (January 2009 and 2013 SSWs studied), the evolution of the ACC and CRPS skill metrics in temperature is next provided for the ensemble-mean of the interactive forecast, together with the difference with respect to control forecast ensemble-mean. To help elucidate the overriding signal, the evolution is shown for the aggregated periods: 0-20, 20-40, and 40-60 days after the event onset, as a function of both latitude and pressure globally during these times. Comparison of the ACC and CRPS skill metrics helps to diagnose the skill change in terms of variability or pattern correlations apart from a change in the mean bias respectively, resulting from inclusion of radiatively interactive HLO.

4.4.1. Composite Approach

4.4.1.1. ACC

As explained in section 2.4, the ACC is a measure of forecast skill in terms of the correlation between two spatial fields of anomalies. The multi-day average ACC in temperature, for three 20-day intervals following the composited SSW onset, is shown in Figure 16. In the high-latitude northern hemisphere stratosphere ($\sim 40\text{-}80^\circ\text{N}$), the ACC is greatest in the first 20 days (panel a), with values as high as 0.85.

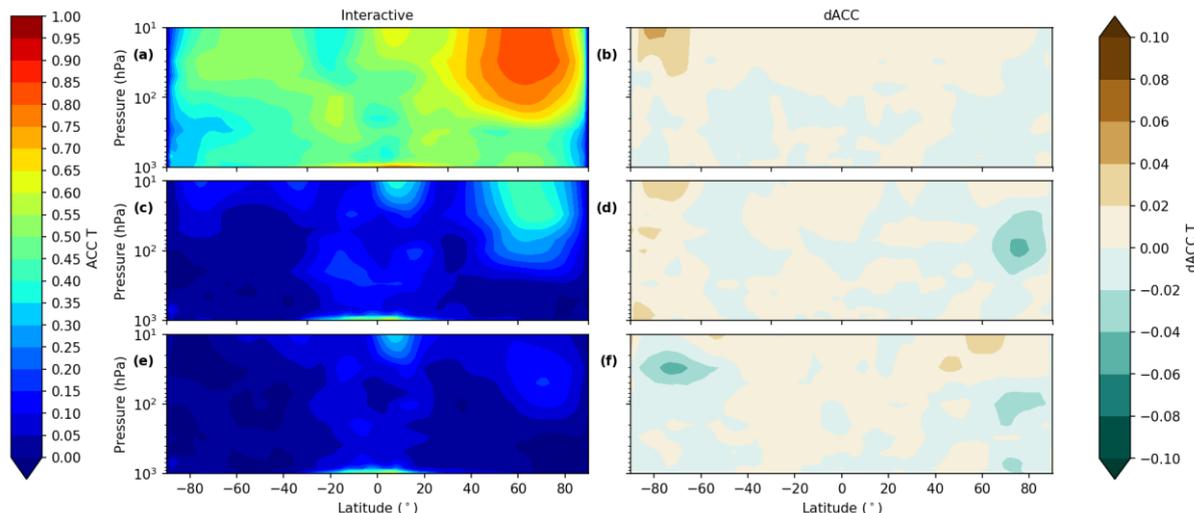


Figure 16: Latitude-pressure resolved temperature ACC for (a) the interactive forecast and (b) the interactive minus control forecast difference (dACC), averaged 0-20 days after the composited SSW onset date (when each set of forecasts were initialised). The same information is subsequently shown in (c) and (d) averaged 20-40 days after the event onset and again in (e) and (f) averaged 40-60 days after the event onset. The ACC was computed over all longitude points for each latitude ($n = 91$) and for all 51 ensemble members using the WMO equation (see section 2.3), which accounts for the latitudinally resolved mean bias.

For context, a value in ACC above 0.6 is considered to represent skilful prediction of spatial anomalies at a synoptic-scale level (e.g. Krishnamurti et al., 2003). This significant source of predictability is striking when compared with an ACC value of 0.4 to 0.6 more typically elsewhere in the stratosphere and troposphere over this period, as highlighted earlier in section 1 (Figure 1). Thus, skilful prediction is marginal at best at the synoptic level, away from the high-latitude northern hemisphere stratosphere, when averaged over the first 20 days as a whole. The difference in ACC (dACC) between the ensemble-mean of the interactive and control forecasts (panel b) shows a very slight overall improvement in stratospheric skill, in response to implementing radiatively interactive HLO (represented by lightest brown shading of up to 0.02 or 2%). A stronger and more coherent signal is evident over the high-latitude southern hemisphere (during summertime) between 10 and 50 hPa, where the ACC is approximately 2-6% higher. This highlights again the importance of solar heating in enabling a benefit from radiatively interactive HLO to manifest in skill scores such as ACC (earlier demonstrated in section 3.1), which is not replicated in the northern hemisphere polar region despite anomalously high ozone following the SSW.

For the subsequent 20-40 day period, the average temperature ACC (panel c) falls markedly (values below 0.3 widespread), which implies that predictability is weak or lost by this time range. The exception is between 50 and 80°N (10-70 hPa) where the ACC stays as high as 0.5, reflecting the long memory of the SSW signal. Although the ACC is close to 0 beneath this region (in the troposphere), the importance of the stratosphere for predictability on sub-seasonal timescales is clearly exemplified for this event. The overall signal of the dACC during time (panel d) is inconclusive across most latitudes. A continued signal for improvement in the spatial representation of stratospheric temperature anomalies persists over the southern hemisphere polar-cap region (60-90°S), of up to around 5%, which also extends into the troposphere during this period. In contrast, a signal for degradation emerges over the Arctic (60-90°N) region due to the HLO scheme, primarily between 30 and 200 hPa (also up to 5%), which is consistent with the region characterised by the LMS cold bias (Polichtchouk et al., 2021).

The ACC in temperature averaged between 40 and 60 days following the event is largely unchanged (panel e), except for the northern hemisphere high-latitude stratosphere where the ACC drops to low values (~0.25). Nonetheless, a small degree of skill persistence remains out to this lead time as values are still elevated compared with the rest of the extratropical stratosphere. The dACC field (panel f) instead shows a reduction in temperature ACC over the Antarctic region, most evident in the mid-stratosphere (~20-50 hPa) by up to 5%, which is also applicable over the Arctic lower stratosphere and troposphere region. Interestingly however, a slight positive signal emerges towards the mid-stratosphere here (>2%) which is perhaps a response to increasing solar insolation with the seasonal transition from winter to spring.

4.4.1.2. CRPS

The CRPS is another important verification metric that may be applied to discern skill change between two sets of ensemble forecasts, although it differs to the ACC skill metric in that it is also sensitive to any change to the mean bias and it is negatively orientated (see section 2.4 for more information). The same three 20-day periods are shown for the composite of all interactive experiments, as well as the difference (dCRPS) with respect all control experiments, in Figure 17. Indicative of increased forecast temperature error, the CRPS tends to be greater in the extratropical troposphere for each hemisphere, as well as the high-latitude northern hemisphere stratosphere also (values of up to $\sim 2^\circ\text{C}$), during the 0-20 day range (panel a). This is expected as the troposphere is inherently more chaotic, and the latter can be explained by the large changes in temperature occurring in conjunction with the SSW. The dCRPS field (panel b) highlights little skill change in these regions, but a general deterioration in the middle to lower stratosphere over the tropics and southern hemisphere (of $>15\%$). As such feature was not apparent for the dACC (Figure 16b), it is inferred that such loss of skill due to the radiatively interactive HLO scheme arises from a degradation of the mean bias (although this warrants verification). It is worth remembering that a $\sim 15\%$ or slightly greater increase in CRPS is however small in absolute terms, as values are characteristically low for these regions. Furthermore, an apparent degradation due to the HLO scheme likely not reflects poor performance, but rather the issue in harnessing additional skill when other modelling deficiencies in IFS exist.

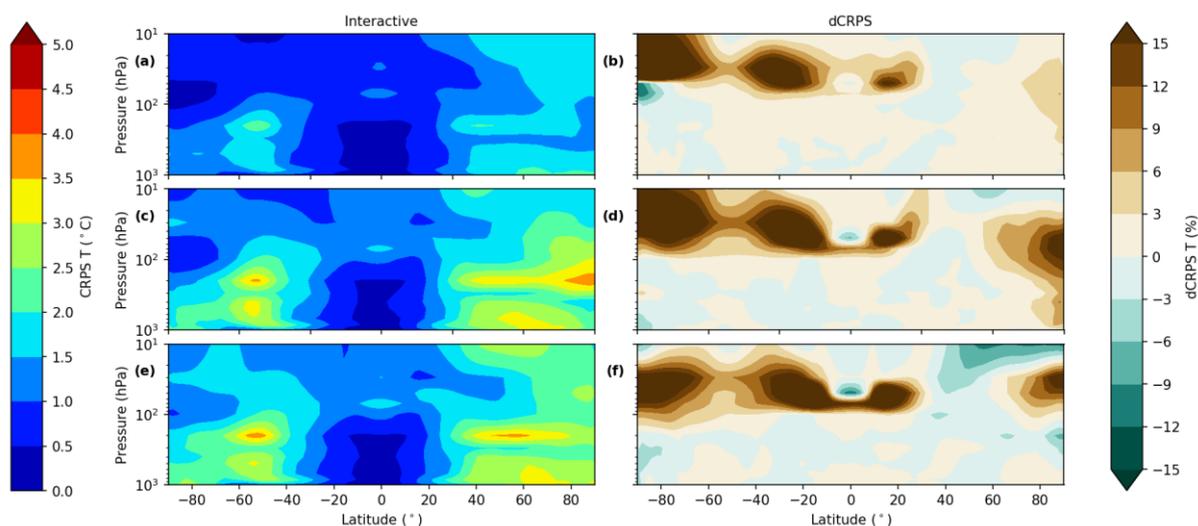


Figure 17: Latitude-pressure resolved temperature CRPS for (a) the interactive forecast, (b) the interactive minus control forecast difference (dCRPS), averaged 0-20 days after the composited SSW onset date (when each set of forecasts were initialised). The same information is subsequently shown in (c) and (d) averaged 20-40 days after the event onset and again in (e) and (f) averaged 40-60 days after the event onset. The CRPS was computed over all longitude grid points ($n = 91$) for all 51 ensemble members.

The 20-40 day average CRPS (panel c) shows an overall very similar theme, albeit with minimal increase in the tropics and stratosphere, and a much more pronounced increase for the extratropical troposphere and UTLS in each hemisphere (increasing to as much as 3-4°C). The dCRPS for this time (panel d) shows a remarkably similar picture to the first 20 days, with degradation affecting much of the stratosphere (>15%), but particularly over the tropics and summertime southern hemisphere. A major difference, however, is the deterioration in the CRPS skill metric for temperature over the Arctic region (up to 15% near the pole between 30 and 200 hPa), which is entirely consistent with the ACC degradation signal during this time (Figure 16d). This implies that the HLO scheme degrades both the mean temperature bias and variability aspect of skill during this forecast range here. Although the dCRPS is small throughout the troposphere and LMS region, there is if anything a signal for slight improvement, which also emerges in the high-latitude northern hemisphere mid-stratosphere between 20 and 40 days following the SSW onset.

For the 40-60 day period, the CRPS field (panel e) again changes little. Values of between 3 and 4°C remain over the extratropical troposphere and LMS, with the lowest values at this time within the tropics (<1°C in the troposphere). The CRPS however increases further (up to 1.5-2.0°C) in the extratropical stratosphere, compared to the earlier 20-40 day period. The dACC (panel f) is largely unchanged from before, apart from the Arctic stratosphere. The degradation signal noted in the earlier 20-40 day period is more localised (still up to 15% between 30 and 100 hPa spanning 70-90°N only), with an opposing signal for improvement above 30 hPa (up to 10%) here. Both features are again consistent with the dACC signal (Figure 16f). This positive signal extends down into the mid-latitude stratosphere and spans poleward again in the LMS (~100-300 hPa), albeit weaker in magnitude (up to 5%). The sensitivity of the CRPS metric to change in the mean bias suggests that the improved temperature skill manifests not only in terms of pattern correlation, but also a reduction in mean forecast error over the Arctic (60-90°N) region. This is however not applicable over mid-latitudes and in the polar LMS, where such positive impact inferred from the CRPS metric is not reciprocated using the ACC skill score (suggesting mean temperature bias is improved but not pattern correlation here). Nonetheless, it is worth noting that predictability during this forecast range is likely to be very limited, and the robustness of such skill changes may still be questionable despite compositing over all events. Variance in the signals for individual events is next illustrated for both the January 2009 and 2013 SSW events.

4.4.2. 24 January 2009

4.4.2.1. ACC

The ensemble-mean temperature ACC for the interactive experiment, together with the difference (dACC) compared to the control experiment, is next shown in Figure 18 (averaged for the same three 20-day intervals following the 2009 SSW onset). The latitude versus pressure resolved ACC distribution is consistent for each period (panels a,c,e), relative to that shown earlier for the composite approach

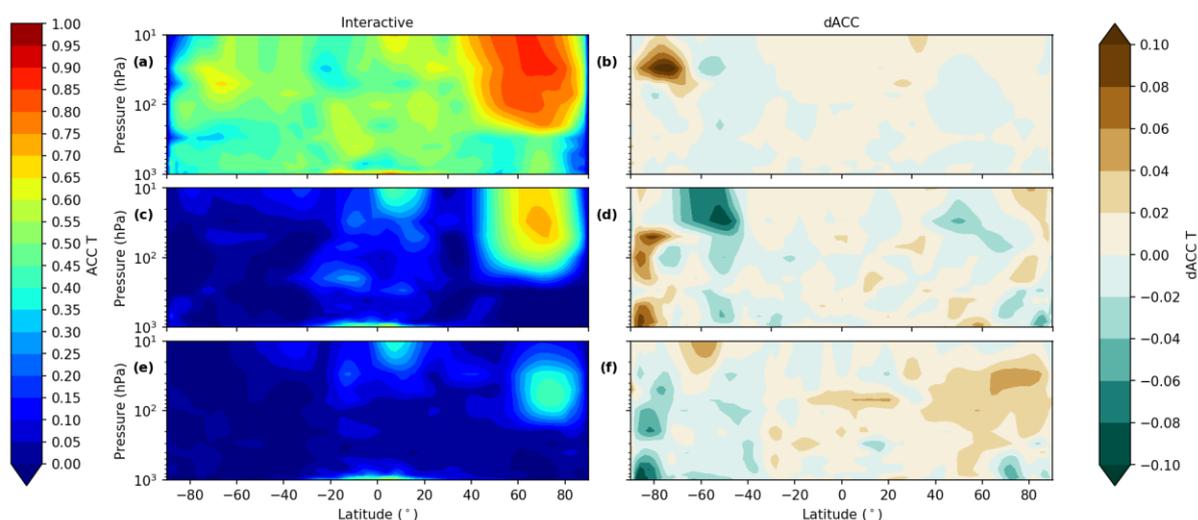


Figure 18: Same as Figure 16 but for the 24th of January 2009 SSW event only.

(Figure 16a,c,e). The high-latitude northern hemisphere SSW signal is however stronger and more prolonged compared with the average evolution over all events, extending deeper into the lower stratosphere also (as apparent from the sharp gradient in ACC extending down to ~200 hPa in both the 0-20 and 20-40 day forecast range). The temperature ACC here remains as high as 0.7 between 20 and 40 days and 0.4 between 40 and 60 days after the event onset (which compared with ~0.5 and 0.25 for the composite approach).

The pattern in the dACC for each 20-day period differs to a larger degree compared with the composite evolution (panels b,d,f). The overall signal is still similar in the 0-20 day and 20-40 day range, with clear indication of improvement in the high-latitude southern hemisphere stratosphere in the first 20 days (>10% around 30 hPa between 60 and 90°S), extending into the troposphere over the subsequent 20 days. An overall negative residual in ACC is too shown between 20 and 40 days in the high-latitude northern hemisphere stratosphere once more, although the signal is patchier than was shown in the composite case (Figure 16d). However, a degradation in temperature ACC emerges in the southern hemisphere between 50 and 70°S (particularly in the stratosphere), which was not so evident when the evolution was aggregated over all events during this time. The largest differences in the dACC are unsurprisingly apparent for the 40-60 day forecast range. Although a negative dACC signal is present for both this event and the composite case over the southern hemisphere extratropics, the signal is one of improvement widely in the northern hemisphere (up to ~5% between 30 and 200 hPa, poleward of 40°N). This event is interesting as it was unusually intense and deeply propagating (Manney et al., 2009a), so it is logical that radiatively interactive HLO might have a more pronounced positive impact following this event compared to the positive case. Indeed, it was shown in section 4 that the magnitude

and persistence of positive polar-cap (60-90°N) anomalies in ozone was greater for the January 2009 event, with respect to the composite or January 2013 event, particularly in the LMS.

4.4.2.2. CRPS

Averaged again over the same three 20-day periods following the January 2009 SSW, Figure 18 shows the interactive temperature ACC for the ensemble-mean (panels a,c,e), as well as the difference (dCRPS) with respect to the control experiment (panels b,d,f). The interactive CRPS evolution matches closely that for the composite case as expected (any skill change due to HLO is always a fraction of that of the total forecast error). Nonetheless, the dCRPS following the 2009 event is still very similar for both the 0-20 and 20-40 day forecast range (albeit the signal is more spatially variable as would be anticipated for a reduced sample size). The only notable difference emerges between 40 and 60 days after the event onset in the extratropical northern hemisphere. Following the January 2009 event, the dCRPS is overall consistent with the dACC (Figure 18f), in that the HLO scheme appears to improve the temperature representation, principally within the polar-cap (60-90°N) region by up to 15%.

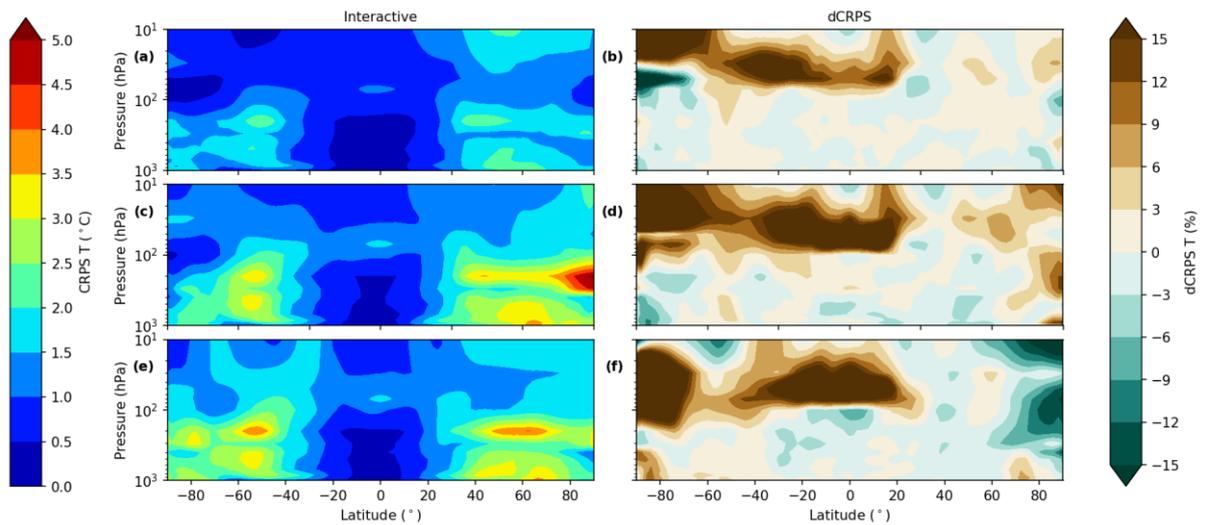


Figure 19: Same as Figure 17 but for the 24th of January 2009 SSW event only.

4.4.3. 24 January 2013

To compare and contrast signals in the performance of the HLO scheme in regard to temperature predictability following the January 2009 SSW event, with the aid of the ACC and CRPS skill metrics, the assessment is now extended to cover the 60 days following the January 2013 SSW.

4.4.3.1. ACC

The multi-day average variability in the temperature ACC, as a function of pressure and latitude, is further shown in Figure 20 (for the periods 0-20, 20-40 and 40-60 days) after the onset of the January 2013 SSW. This particular event was less pronounced than the January 2009 event, which is reflected by the temperature ACC signal in the extratropical northern hemisphere stratosphere (panels a,c,e). Though the values are generally higher still than for the composite case after the first 20 days (Figure 16), the vertical extent of high ACC values (associated with the long memory of the SSW signal) are largely confined to the region above 70 hPa after 20 days for the January 2013 SSW event. The patterns in dACC (panels b,d,f) vary quite substantially compared to the January 2009 event. The main commonality is the enhancement in temperature ACC for the summertime polar southern hemisphere stratosphere region (60-90°S) between 0 and 20 days (>10% near 30 hPa). For this particular event, the strength of this signal however not only extends into the 20-40 day range, but also intensifies and becomes more expansive. This was not so apparent for the January 2009 event (Figure 18c) but does agree more closely with that shown for the composite case (Figure 16c). The dACC reflects once more a degradation in the 40-60 day range however, spanning much of the extratropics (widely between 5 and 10% for this event in both the stratosphere and troposphere). In the northern hemisphere, a negative residual in ACC is overall widespread between 20 and 40 days following the event onset (~2-4%), except over the Arctic (between 10 and 100 hPa) where a signal for improvement due to HLO,

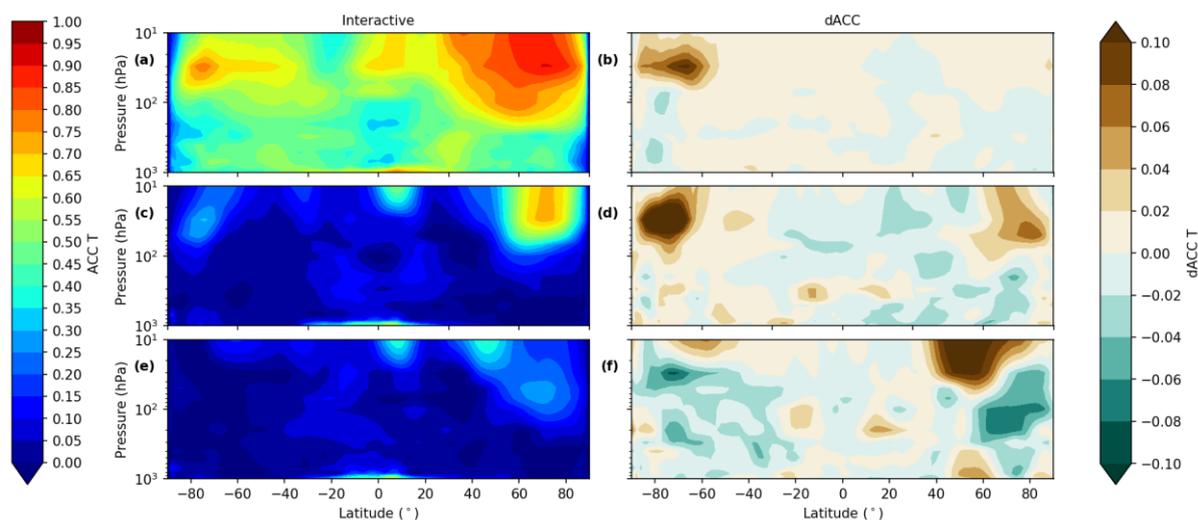


Figure 20: Same as Figure 16 and 18 but for the 7th of January 2013 SSW event.

interactive with radiation, emerges (up to 5%). The latter feature in particular is less supported by the composite case, so appears to be an event-specific attribute. The dACC pattern between 40 and 60 days (panel f) conforms with that shown for the composite case (Figure 16f), unlike that of the January 2009 SSW event (Figure 18f), although the magnitude is significantly larger for this event. The result is that the spatial representation of temperature in the northern hemisphere extratropics is improved by greater than 10% between 10 and 30 hPa, but degraded by up to 6-8% in the polar LMS. The degradation signal here due to radiatively interactive HLO may not be robust but could be influenced by the premature decay of the characteristic positive lower stratospheric ozone anomaly following the SSW, which was much more in evidence during this time according to the CAMS reanalysis (Figure 15).

4.4.3.2. CRPS

The corresponding evolution in the ensemble-mean temperature CRPS following the January 2013 SSW is subsequently shown in Figure 21 for the interactive experiment, together with the normalised difference (dCRPS) with respect to the control forecast ensemble-mean. The overall evolution in the CRPS values (panels a,c,e) is again very similar to the composite and January 2009 SSW case. The dCRPS evolution (panels b,d,f) is again quite similar, particularly in the southern hemisphere and tropics where a degradation to the stratospheric mean temperature bias due to radiatively interactive HLO is the prevailing feature across all three 20-day periods (widely >15%). One localised exception is in the tropics around 50 hPa where HLO, interactive with radiation, appears to improve a region of relatively poor skill (>1°C but reduced by as much as 15% in the interactive case). This is likely associated with the QBO, which was in an easterly phase during the winter of 2012/13, although such signal was curiously not seen for the 2009 event when the QBO was in its westerly phase (Lu et al., 2020). Despite this, the operational use of the HLO scheme in the System 4 seasonal forecasting system was shown to improve QBO representation, so such feature is perhaps not unexpected. Indeed, any such

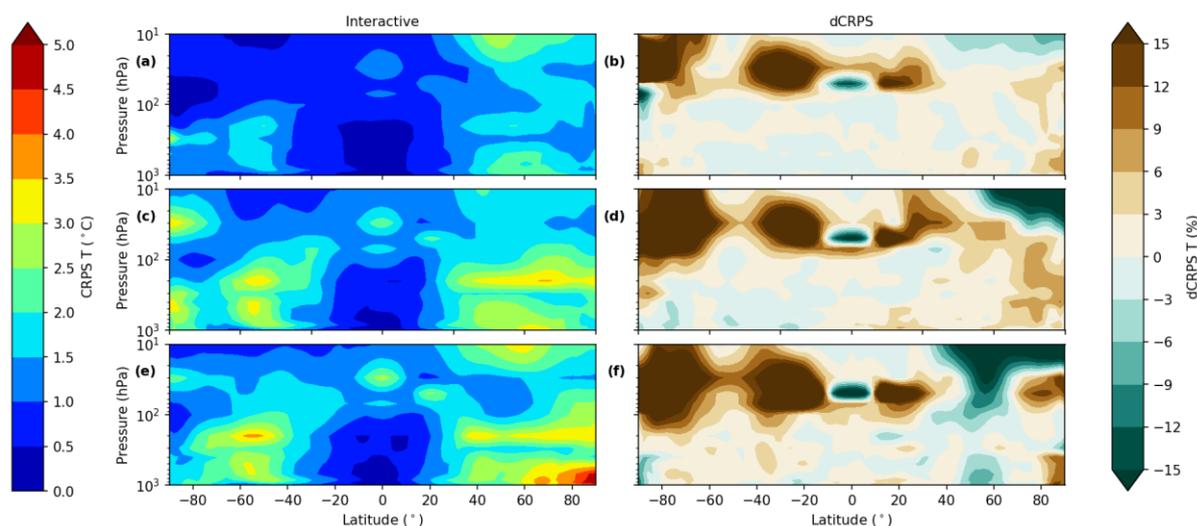


Figure 21: Same as Figure 17 and 19 but for the 7th of January 2013 SSW event.

amelioration of the QBO representation due to HLO, interactive with radiation, would be of particular interest as it constitutes one of the few pure sources of predictability on seasonal timescales (Ebdon, 1975; Folland et al., 2012). In the northern hemisphere extratropics, the dCRPS field for each period exhibits many localised differences with respect to the composite and January 2009 SSW case, but the overall signal more closely matches that of the former (Figure 17). The region characterised by the LMS cold bias appears to be degraded by implementation of radiatively interactive HLO, except between 40 and 60 days, whilst temperature representation in the mid-stratosphere improves with increasing forecast lead time. The skill implications of implementing radiatively interactive HLO require more in-depth investigation to confidently conclude any impact to the representation of stratosphere-troposphere coupling, of particular interest for sub-seasonal prediction following such events. In an attempt to understand the connection between ozone, radiation and temperature more comprehensively following PJO-type midwinter SSWs, results are subsequently show in section 5 for the January 2009 SSW case study, involving examination of radiative heating tendencies and the attenuation radiatively interactive HLO has on this.

5. Radiative heating tendency investigation – January 2009 SSW case study

To explore the mechanism behind the impact of the radiatively interactive HLO scheme on temperature forecast skill within IFS, shown earlier in sections 3 and 4 for both the medium and sub-seasonal forecasting range respectively, the evolution of stratospheric radiative heating tendencies following the major northern hemisphere January 2009 SSW event is next investigated. As indicated in Table 1 and section 2.5, both a control and interactive 60-day, 51-member ensemble experiment was initialised on the 24th of January 2009 (SSW central warming date). The relationship between the evolution in ozone, temperature and radiative heating tendencies is first examined over the southern hemisphere for the same forecast window, encompassing the quiescent state of the summertime southern hemisphere when the solar input is greater. The interpretations made here help to understand the radiative impact of HLO in the forecast over the northern hemisphere, which is complicated by weaker solar input as well as dynamical influence, particularly just after the SSW onset (early in the forecast). Dynamical activity (variations in polar mean downwelling) can strongly influence the evolution in temperature via modulation to adiabatic heating rates, which is a dominant influence around the timing of an SSW onset, accounting for most of the warming associated with these events. However, as winds reverse from westerly to easterly, gradually propagating downwards through the stratosphere, this influence dampens significantly, and stratospheric temperatures are subject to a dominant radiative control. The presence of long-lived ozone anomalies following such events has been recently shown to affect the temperature of the radiatively sensitive UTLS region by as much as $\sim 1^\circ\text{C}$ (Haase and Matthes, 2019; Williams et al., 2021, in prep.). This small effect may be significant in the absence of dynamical induced temperature variations, with possible implications for predictive skill of stratosphere-troposphere coupling mechanisms.

5.1. Interactions between ozone, temperature and radiation

To showcase the relationship between spatial anomalies in ozone, temperature (as quantified by the difference between the ensemble-mean of the interactive and control experiments) and radiation (both absolute and corresponding forecast differences in heating tendencies), a series of snapshot forecast fields in polar stereographic projection are shown for a forecast lead time of 7 days (T+168) in each hemisphere. To illustrate these linkages in both the middle and lower stratosphere, pressure levels ~ 10 hPa and ~ 70 hPa (nearest IFS model levels extracted) are here examined respectively. The short or medium range is focussed on here as forecast non-linearities become sufficiently large to complicate any interpretations made when evaluating at longer lead times.

5.1.1. Southern hemisphere

7-day (T+168) ozone and temperature anomaly forecast fields, with respect to a monthly-mean climatology from ERA-5, are first shown for 10 hPa over the southern hemisphere in Figure 22 (for 31st January 2009), together with the ensemble-mean difference (interactive minus control) in projected temperature and the respective difference against the ERA-5 analysis for both variables. In Figure 23, the corresponding radiative heating tendencies (in units of K day⁻¹) are shown for the interactive experiment (LW, SW and Net), in addition to the ensemble-mean of the (interactive minus control) difference in these components. The spatial pattern in projected ozone and temperature anomalies in

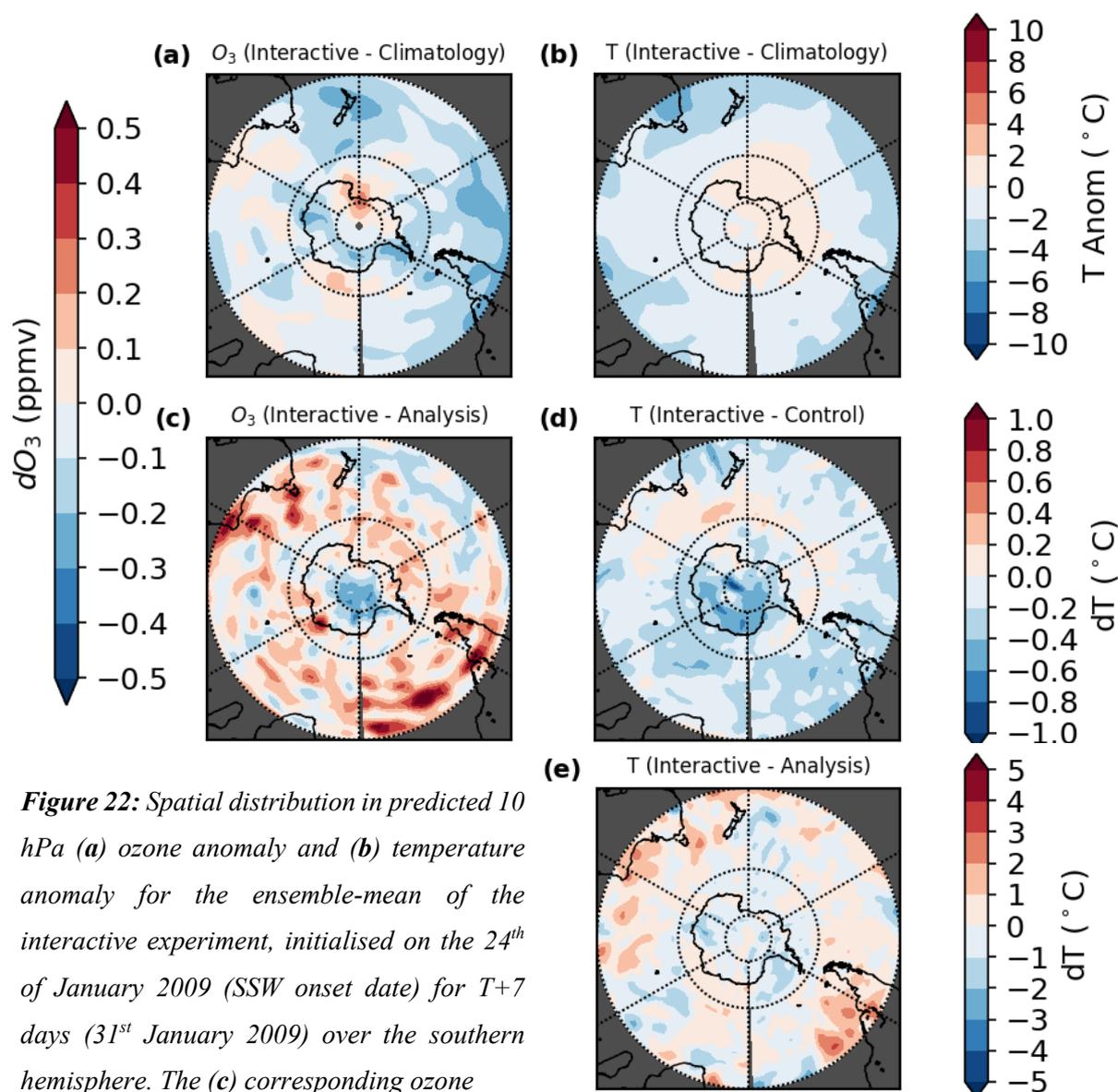


Figure 22: Spatial distribution in predicted 10 hPa (a) ozone anomaly and (b) temperature anomaly for the ensemble-mean of the interactive experiment, initialised on the 24th of January 2009 (SSW onset date) for T+7 days (31st January 2009) over the southern hemisphere. The (c) corresponding ozone difference with respect to ERA-5 analysis; (d) ensemble-mean interactive minus control temperature difference (ΔT) and (e) interactive temperature versus ERA-5 difference is additionally shown.

this instance (for the interactive experiment) appears to be weakly correlated (Figure 22a-b), although it should be noted that spatial variations are rather muted at this level during late austral summer. With respect to ERA-5, the ozone difference field (Figure 22c) shows much greater spatial complexity, with an overall tendency for the HLO scheme to overestimate the ozone volume mixing ratio (ppmv) at this level. Nonetheless, the anomaly in interactive ozone is generally negative on this particular day, which results in overall reduced SW heating (shown next in Figure 23e) compared with the control. This likely explains the general cooling impact of the radiatively interactive HLO scheme (up to -1°C) across the southern hemisphere ($30\text{-}90^{\circ}\text{N}$), as evidenced in Figure 22d. In terms of the difference in temperature (simulated by the interactive experiment) relative to ERA-5 (Figure 22e), a slight overall cold bias is evident over the Antarctic continent, with an overall warm bias near southern Australia and southernmost South America (up to 2°C in each case). These regions tend to coincide with a negative and positive residual in the projected ozone mixing ratio respectively (consistent with local radiative control). Whilst it is important to realise that the difference in ensemble-mean interactive temperature, with respect to the ERA-5, is typically much larger than when compared against the control forecast, the impact of radiatively interactive HLO is again positive in this instance (in that temperature forecast bias is overall reduced). Despite the dynamically inert state of the summer stratosphere, temperature forecast errors may manifest from other modelling deficiencies (e.g. parametrization of nonorographic gravity wave drag; see Polichtchouk et al., 2018 for instance) other than the radiative effects of ozone, which can explain such disparity in the magnitude of the interactive minus control and interactive minus ERA-5 differences.

The daily-mean heating tendency for the interactive forecast (Figure 23a-c), between forecast day 6 and 7, shows a very uniform situation over the southern hemisphere extratropics ($30\text{-}90^{\circ}\text{S}$). An overall slight warming prevails (up to $\sim 1\text{K day}^{-1}$ near 30°S), except over and around Antarctic where the net heating tendency is negative ($< 0\text{K day}^{-1}$). The spatial pattern of the net tendency reflects the balance between SW heating and LW cooling (of ~ 3 to 4 K day^{-1}), which largely offset each other. The difference in these radiative components (interactive minus control experiment) help to demonstrate the prevailing impact of the radiatively interactive HLO scheme (Figure 23d-f). A positive residual in LW heating translates to reduced cooling, whilst a negative difference in SW heating reflects reduced heating ($\sim \pm 0.1\text{-}0.25\text{ K day}^{-1}$). The Net heating tendency difference suggestive of a warming effect over the Southern Ocean (between 60°W and 60°E) and around Australia/New Zealand and a cooling effect elsewhere ($\sim 0.1\text{ K day}^{-1}$).

Ozone and temperature anomalies in the lower stratosphere ($\sim 70\text{ hPa}$) for the same forecast time (Figure 24a-b) on the other hand exhibit much greater hemispheric variability. This is likely related to influence from tropospheric planetary wave activity, which can still propagate into the lower stratosphere in summer, giving rise to coherent regions of positive and negative anomalies. In support of a strong local

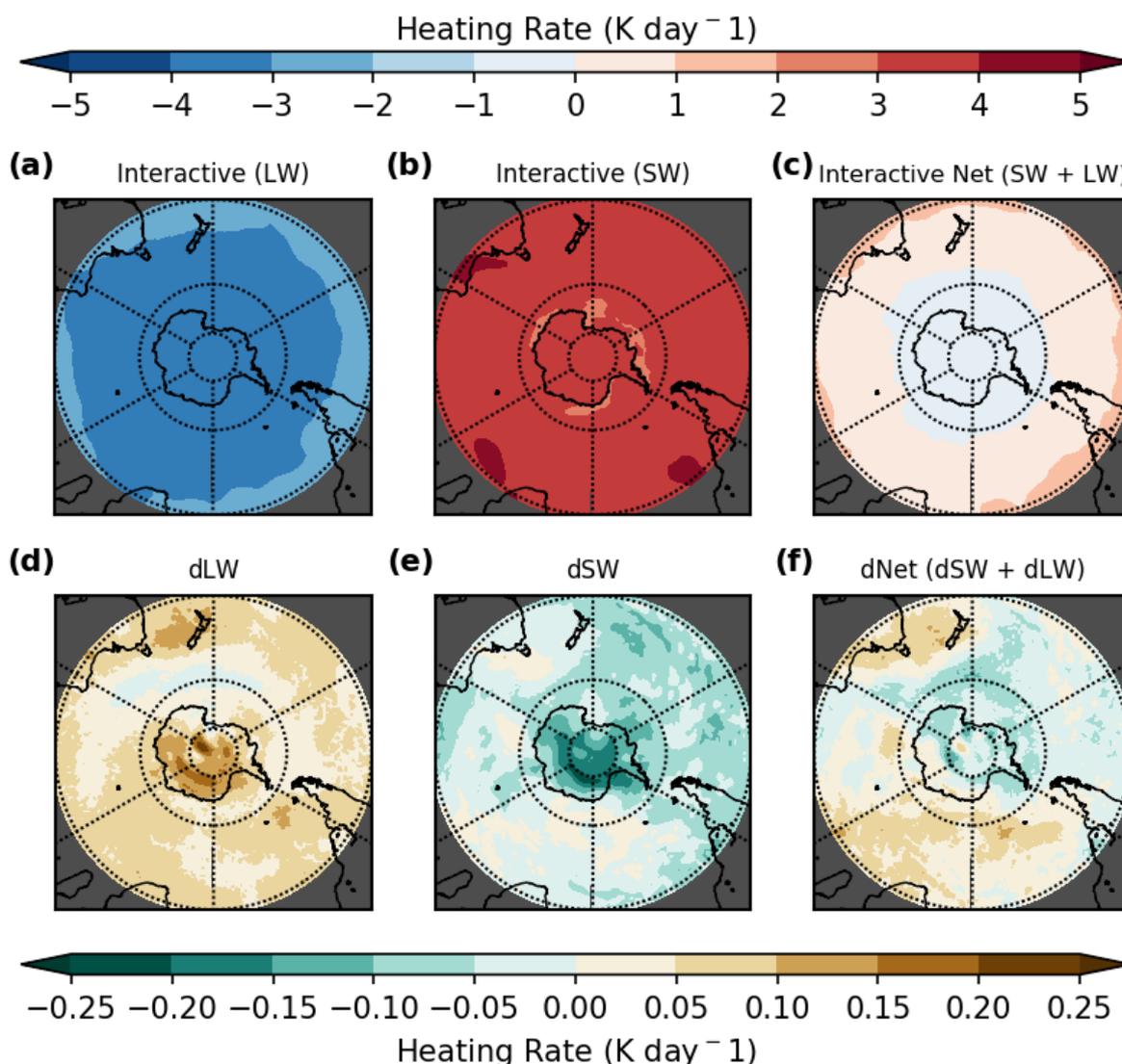


Figure 23: The spatial distribution in the model derived (a) LW heating, (b) SW heating and (c) Net (SW + LW) heating tendency between forecast day 6 and 7 at 10 hPa for the ensemble-mean of the interactive experiment, in addition to the heating tendency difference between the ensemble-mean of the interactive and control forecasts: (d) dLW, (e) dSW and (f) dNet (dSW + dLW) heating.

ozone-temperature feedback at this pressure level, anomalies in ozone and temperature are very closely correlated, with positive (negative) anomalies in ozone coincident with positive (negative) anomalies in temperature. The magnitude of each anomaly is largely proportional, which implies that ozone variability has a highly linear impact over temperature at this level and forecast lead time. This spatial variability imprints onto the interactive minus control forecast difference in temperature (Figure 24d), but the radiative impact of HLO is generally that of a cooling effect ($>0.4^{\circ}\text{C}$ widely north of 60°S). The implication is that the abundance of ozone for the zonal-mean, monthly-mean climatological ozone field, as seen by the radiation scheme for the control experiment, is on average higher than projected

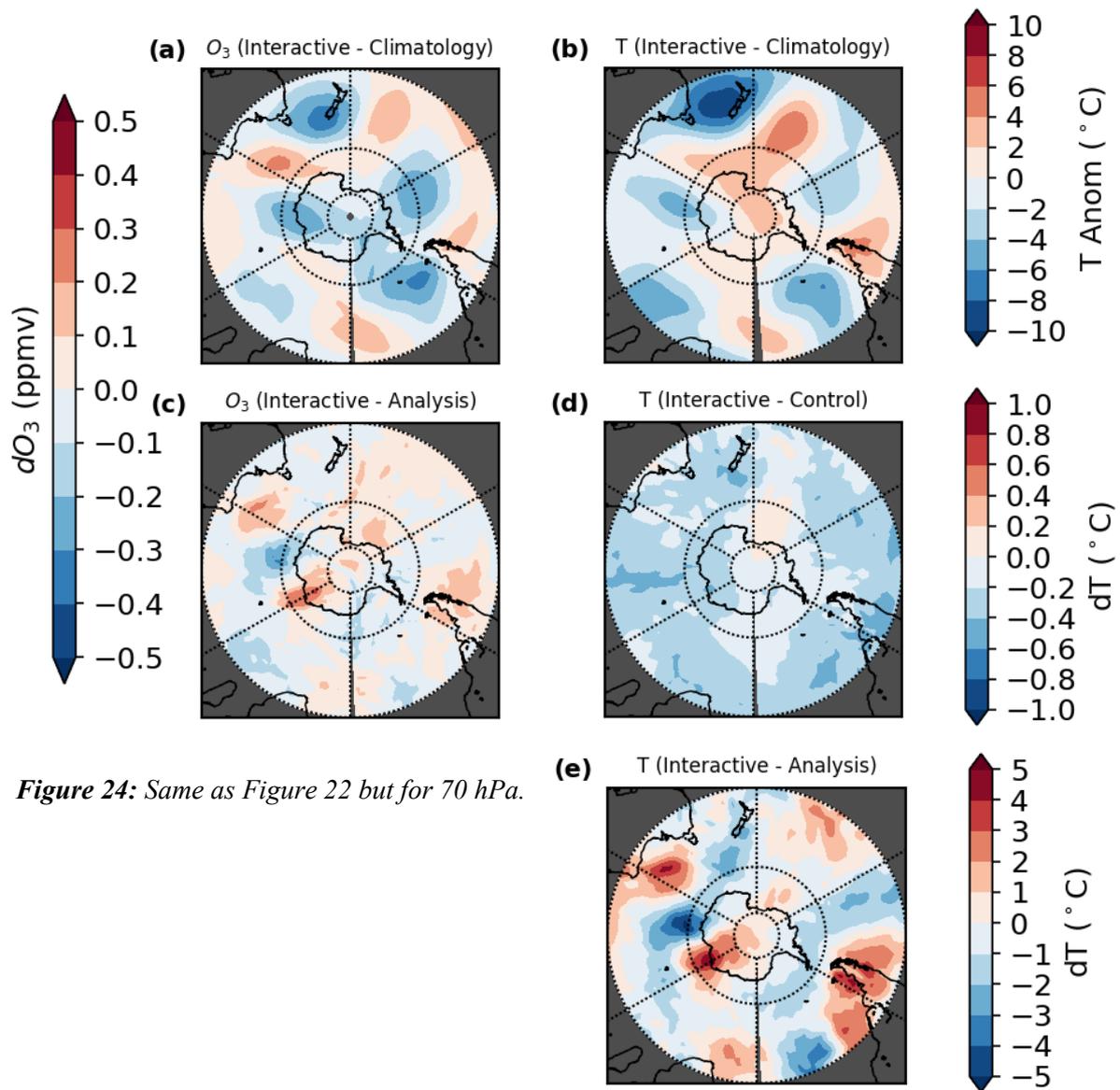


Figure 24: Same as Figure 22 but for 70 hPa.

using the HLO scheme (as would be consistent with the strong local radiative feedback of ozone at this level during summer). Verification against ERA-5 analysis for ozone (Figure 24c) shows that much of HLO-simulated spatial variability in ozone is indeed largely accurate representation. An exception to this might be over parts of the Southern Ocean between east Antarctica and Australia (~60°E-120°E), where differences exceed ~0.2 ppmv. The inaccuracies here translate to same sign temperature forecast errors (Figure 24e), which again reflects the local importance of ozone heating. Errors in projected temperature elsewhere mirror the distribution of interactive ozone closely, but are typically larger than would be anticipated under linear radiative control. At a forecast lead time of 7 days (T+168) however, it is inevitable that phase shifts in the positioning of such features arise, which likely contributes to the spatial pattern of temperature forecast error. Indeed, relatively small forecast errors in ozone (<0.2 ppmv), as seen in Figure 24b, tend to be spatially offset compared with spatial errors in forecast temperature, which supports this assertion.

The corresponding radiative heating tendency spatial distribution (Figure 25) for the 70 hPa level shows an overall slight cooling (warming) response due to HLO, north (south) of $\sim 50^{\circ}\text{S}$. The spatial pattern of the forecast differences (interactive minus control) in the heating tendency components exhibit structure consistent with the interpretation of a local radiative effect due to ozone. Most notably, negative anomalies in ozone (as shown in Figure 24a) coincide with a reduction in the SW heating component, which is offset to an extent by a reduction in LW cooling. The LW cooling response is driven by the ambient temperature, as LW emission is almost entirely regulated by well-mixed carbon dioxide (CO_2). Therefore, the LW cooling effect will be weaker for the simulation characterised by colder temperatures. As each forecast evolves, the divergence in temperature arises first in response to differential SW heating (as the ozone field visible to the radiation scheme is very different between the interactive and control case), which is the primary forcing mechanism early in the forecast when differences in temperature are small. At longer lead times, the counteractive LW cooling response

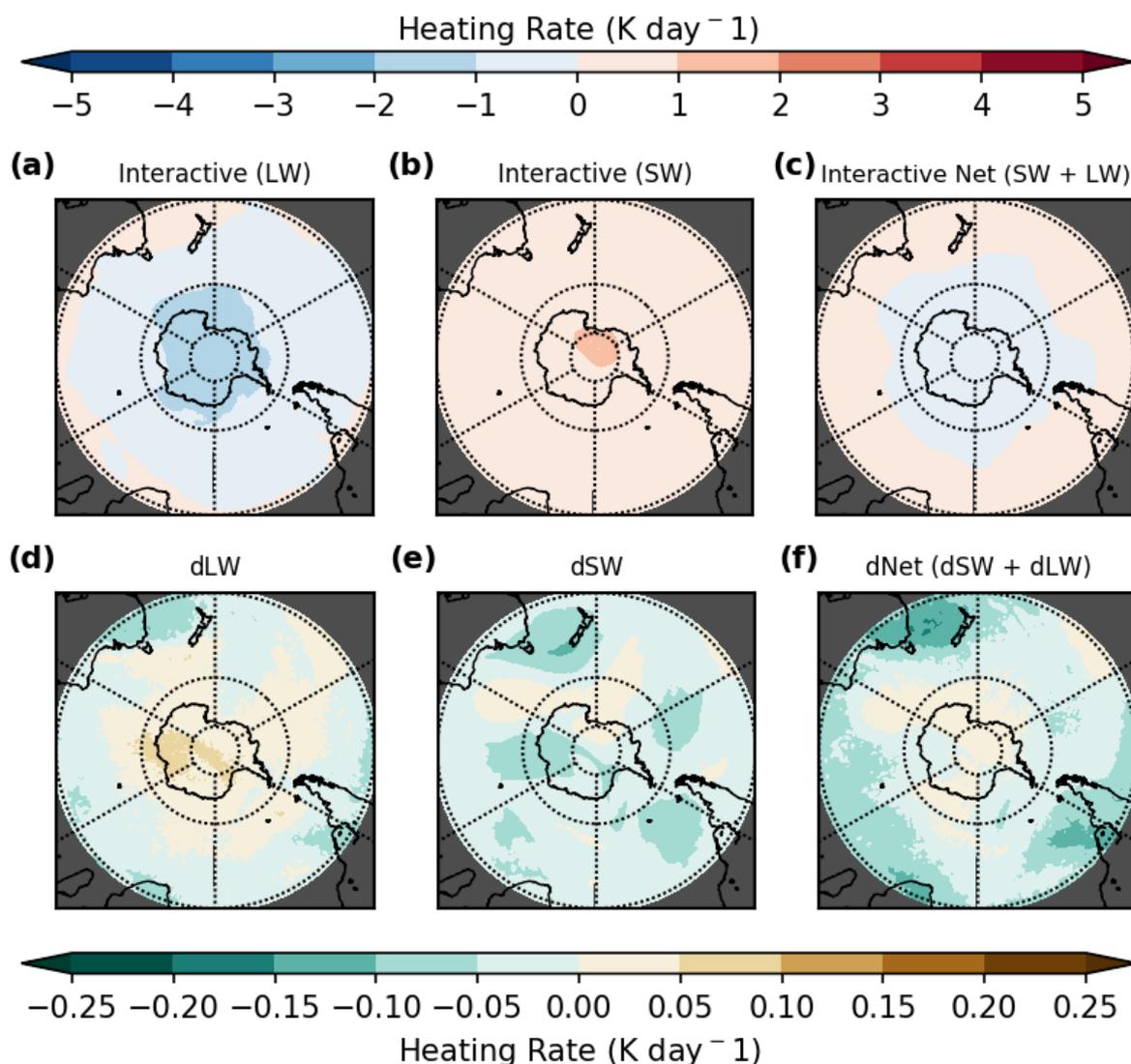


Figure 25: Same as Figure 23 but for 70 hPa.

increasingly acts to offset the modulation to temperature due to SW heating, until a radiative equilibrium is reached. By this stage of the forecast, any correlation between the ozone and net heating rate difference will be weak to negligible. At around 7 days into the forecast, the LW component is shown here to be influential but still smaller than the SW heating response. Thus, the spatial pattern in the net heating difference more closely matches that for the differential SW heating response due to the radiatively interactive HLO scheme.

Subsequently, the SW heating component is generally enhanced where the anomaly in ozone is positive and instead offset to a degree by greater LW cooling. Examination of the net heating tendency difference (Figure 25f) shows that the reduced SW effect dominates, particularly where the anomaly in ozone is negative, which is to be expected given the long day length and relatively high midday position of the sun at this time of the year. The correlation is weaker however due to the LW cooling response to existing anomalies in temperature. This factor indeed complicates the understanding of the local radiative effect of ozone in the experiment comparison performed here, particularly at longer forecast times. Thus, the relationship between ozone, temperature and radiation as a function of lead time is examined later in section 5.2 in an aggregate sense to help better disentangle such competing influences.

5.1.2. Northern hemisphere

To contrast with the quiescent summertime state of the southern hemisphere stratosphere, the 7-day (T+168) forecast snapshot is next shown for the northern hemisphere at the same time, initialised on the 24th of January 2009 (projected for the 31st of January), which immediately follows the onset of the largest and most prolonged midwinter SSW on record (Manney et al., 2009a). As shown in Figure 26a-b for 10 hPa, much of the polar-cap (60-90°N) region is characterised by a projected positive ozone anomaly exceeding 0.5 ppmv, as well as a temperature anomaly greater than 10°C, according to the ensemble-mean of the interactive forecast. Indeed, this also applies to much of the mid-latitudes (30-60°N), albeit with significant zonal asymmetry in the interactive ozone field and a generally negative temperature anomaly at relatively low-latitudes (~30-40°N). This pattern is synchronous with an accelerated BDC during the build-up and onset stage of an SSW, with the sign and magnitude of the anomalies a product of enhanced poleward and downward descent. With respect to the size of the forecast temperature anomaly, the ensemble-mean temperature difference (Figure 26d) between the interactive and control experiment is relatively small and shows comparatively complex spatial structure. Some of the largest differences occur close to the strongest horizontal gradient in ozone and temperature anomaly, including over western North America, the North Atlantic Ocean and eastern Europe (up to $\sim\pm 1^\circ\text{C}$), where streamers of anomalously low and high ozone are in close proximity to each other. This may at first seem surprising given SSWs are associated with the largest deviations in polar ozone relative to climatology (as seen by the radiation scheme for the control experiment), however the temperature anomaly is first and foremost a result of enhanced adiabatic descent

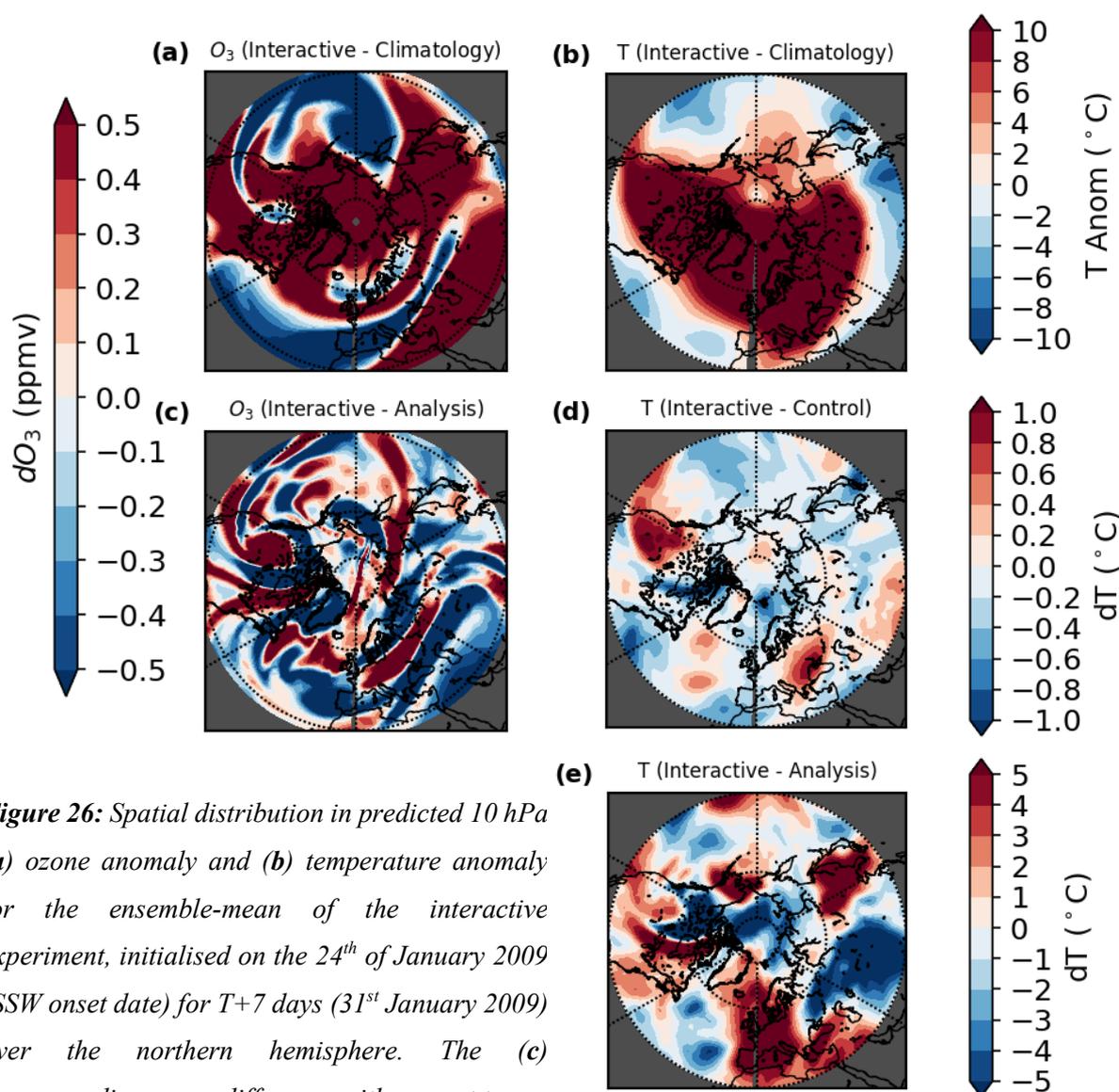


Figure 26: Spatial distribution in predicted 10 hPa (a) ozone anomaly and (b) temperature anomaly for the ensemble-mean of the interactive experiment, initialised on the 24th of January 2009 (SSW onset date) for T+7 days (31st January 2009) over the northern hemisphere. The (c) corresponding ozone difference with respect to ERA-5 analysis; (d) ensemble-mean interactive minus control temperature difference (ΔT) and (e) interactive temperature versus ERA-5 difference is additionally shown.

(dynamical heating). As the solar input is weak at the time of this forecast (negligible at the pole but likely influential over mid-latitudes), it can be inferred that the SW radiative effect of ozone is at least more significant over mid-latitudes.

The interactive minus analysis difference field (Figure 26c), however, shows that much of the HLO projected structure is out of phase with that according to ERA-5 at a 7-day lead time from the SSW central warming date. It should be noted however that this apparent poor performance is sensitive to the extreme conditions associated with the event. Such fine-scale structure (typically less than ~ 200 km) in ozone over mid-latitudes is inherently difficult to predict after a few days, whilst the evolution of stratospheric ozone is usually more predictable. In normal circumstances during northern hemisphere

wintertime, the SPV acts as a transport barrier to meridional transport of ozone from lower latitudes, and thus most regions would usually be characterised by weaker horizontal gradients in ozone volume mixing ratio (with the exception of areas close to this boundary). Despite this, radiatively interactive ozone would be expected to have a larger radiative impact during the extreme SSW case (most different from climatology), however, this influence in the early stage of an event is likely to be significantly outweighed by influence from dynamical heating. The interactive minus analysis difference in temperature (Figure 26e) exhibits coherent spatial structure, with regions characterised by a notable forecast error in temperature ($\pm 5^\circ\text{C}$). Whilst some of it appears to be related to the implementation of radiatively interactive HLO, as evidenced by close spatial correlation with the structure shown in Figure 26a,c,d over regions such as North America in particular, other regions such as Asia and the North Pacific Ocean show no obvious connection. The anti-phase sign of the interactive minus control (Figure 26d) and interactive minus analysis (Figure 26e) difference over the Atlantic and North American sector however lends support to a small improvement in temperature forecast error here. In any case, the temperature difference between the ensemble-mean of each experiment is at least five times smaller than the temperature error with respect to ERA-5.

The corresponding daily-mean radiative heating tendencies (SW, LW and Net) between forecast day 6 and 7 at 10 hPa for the ensemble-mean of the interactive experiment, together with the differences with respect to the control forecast ensemble-mean, are next shown in Figure 27. Whilst the SW heating spatial pattern is fairly uniform and zonally symmetrical (Figure 27b), characterised by a gradually increasing contribution with decreasing latitude ($\sim 2\text{-}3\text{K day}^{-1}$ near 30°N), the LW cooling pattern (Figure 27a) shows interesting structure consistent with the dynamical state of the stratosphere at this level (~ 10 hPa). Immediately following the SSW onset, a dynamical split of the SPV resulted in the relocation of two residual, daughter vortices over North America and western Eurasia (e.g. Tao et al., 2015), which is consistent with the regions characterised by the largest LW cooling tendencies (locally $>5\text{K day}^{-1}$). This pattern imprints onto the Net heating tendency spatial distribution (Figure 27c), with a slight damping effect from the SW heating component (typically $\sim 1\text{-}2\text{K day}^{-1}$ where LW cooling is largest). The distribution of the LW and Net heating tendency is largely a response to the intense dynamical heating associated with the event onset however, and is only modulated to a much smaller degree by the implementation of radiatively interactive HLO, which can be discerned subsequently in Figure 27d-f. With reference to the distribution of the anomaly in ozone at this forecast time (Figure 27a), the SW heating component is clearly enhanced (reduced) where the anomaly is positive (negative), as inferred from the interactive minus control difference in the SW (dSW) in Figure 27e. This is particularly relevant over mid-latitudes where solar heating is more influential (locally exceeding $\pm 0.25\text{K day}^{-1}$). The LW cooling difference (Figure 27d) is evidently largely in anti-phase with the SW heating difference, albeit this compensation is only partial (~ 0.1 to 0.2K day^{-1} typically),

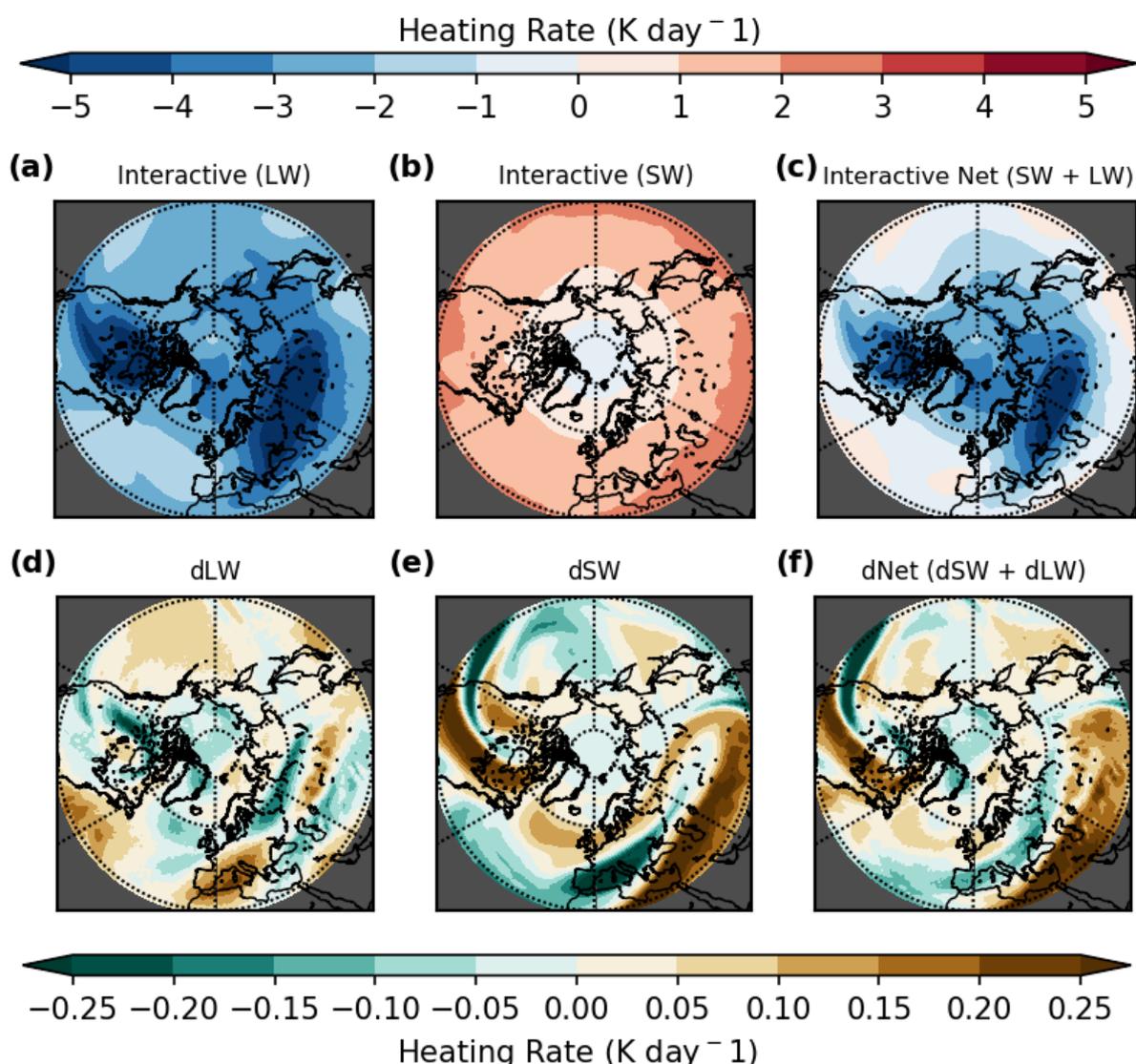


Figure 27: The spatial distribution in the model derived (a) LW heating, (b) SW heating and (c) Net (SW + LW) heating tendency between forecast day 6 and 7 at 10 hPa for the ensemble-mean of the interactive experiment, in addition to the heating tendency difference between the ensemble-mean of the interactive and control experiments: (d) dLW, (e) dSW and (f) dNet (dSW + dLW) heating.

thus resulting in a Net heating tendency difference (dNet) spatial field (Figure 27f) resembling of the dSW spatial pattern. The opposite is true however at high-latitudes (poleward of 60°N) as the solar influence is very small here. The highly perturbed state of the northern hemisphere extratropical mid-stratosphere at this time shows the potential benefit the HLO scheme, interactive with radiation, may bring, but it is again highlighted that this influence is relatively small compared with dynamical activity at this time.

The same forecast snapshot is next displayed for the lower stratosphere (~70 hPa) to compare the radiative impact of HLO as a function of pressure (altitude). The ozone and temperature anomaly spatial distribution are once again shown in Figure 28a-b for 70 hPa at 7-day (T+168) lead time (forecast for the 31st of January 2009) for the ensemble-mean of the interactive forecast. A noteworthy positive anomaly in ozone extends over much of the polar-cap region (60-90°N), largely exceeding 0.5 ppmv, which is similar to that shown for 10 hPa (Figure 26a). In mid-latitudes, the anomaly in ozone is widely negative (up to -0.5 ppmv over some regions such as the US and southwestern Europe) however. This differs to 10 hPa, which had notable streamers of high as well as low ozone, implying that an enhanced downward component of transport prevails in the lower stratosphere. The broad spatial pattern in ozone anomaly closely matches that for temperature, with anomalies exceeding 10°C poleward of 60°N and overall negative anomalies over mid-latitudes (typically between 4 and 8°C). The interactive minus

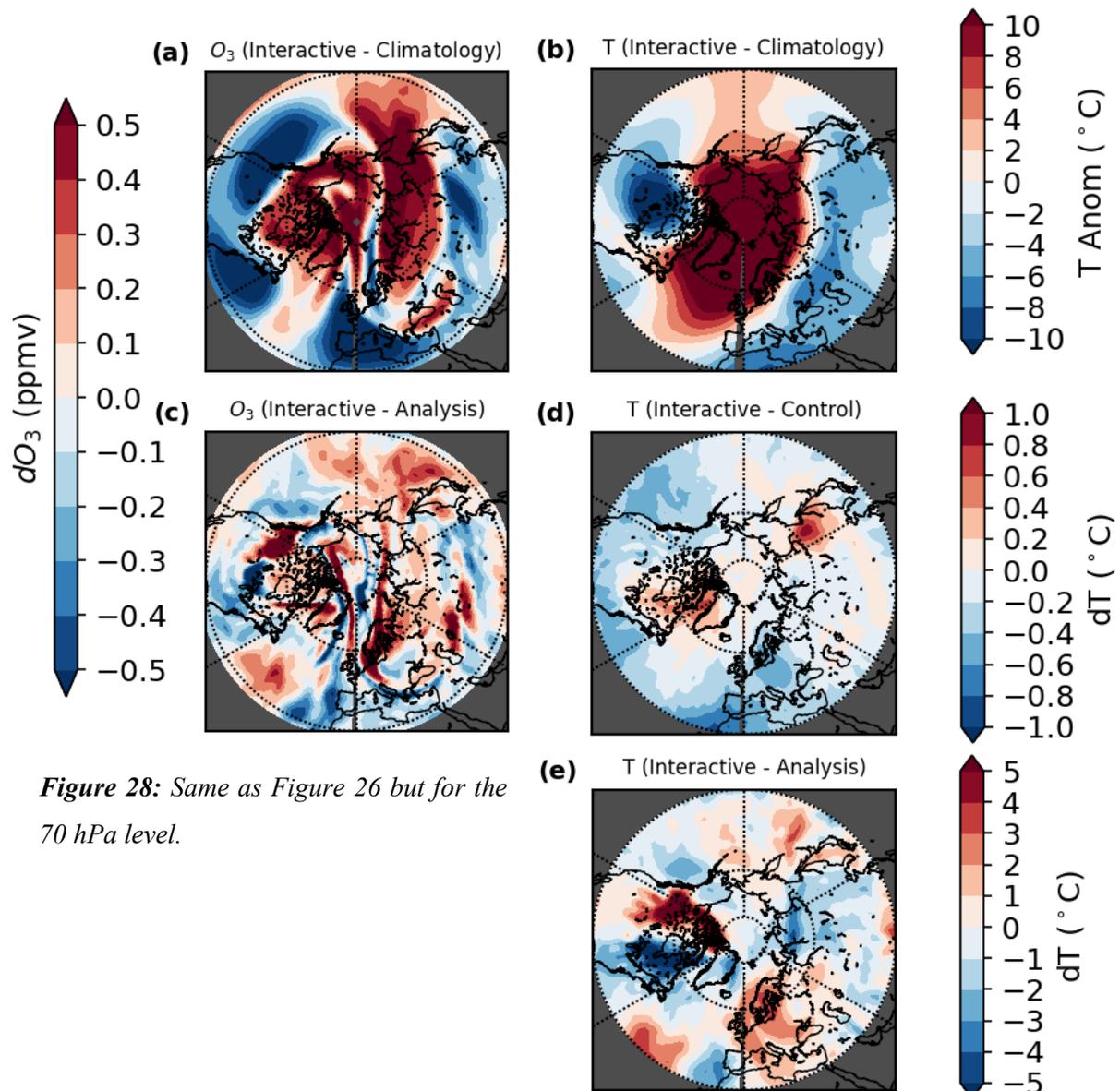


Figure 28: Same as Figure 26 but for the 70 hPa level.

control difference in temperature (Figure 28d) shows an overall slight cooling due to the HLO scheme over the eastern Pacific, US, North Atlantic and much of Europe (~ 0.2 - 0.6°C) and local warming west of Greenland and eastern Russia (up to $\sim 0.5^{\circ}\text{C}$). Similar to that shown for 10 hPa, the difference in interactive ozone with respect to ERA-5 is supportive of a phase shift in fine-scale spatial structure at this forecast range (Figure 28c), manifesting in localised anomalies in ozone (≥ 0.5 ppmv). This is more confined at this level to the polar-cap (60 - 90°N) region, however, and forecast errors in ozone are typically much smaller over mid-latitudes (~ 0.2 to 0.3 ppmv). Again, as noted for the 10 hPa level, the temperature forecast errors for the interactive experiment (Figure 28e) are around five times larger than the respective differences against the control experiment (up to $\sim \pm 5^{\circ}\text{C}$). Nonetheless, an overall anti-phase sign of the differences (consistent with that remarked upon for 10 hPa) implies that the HLO scheme serves to slightly reduce the overall temperature forecast error at this lead time (by ~ 10 - 20%), which again is most apparent over Europe, the North Atlantic, North America and eastern Pacific regions.

The projected radiative heating tendency and interactive minus control forecast difference spatial patterns are additionally shown in Figure 29. Compared to the ~ 10 hPa level (Figure 27), the radiative heating tendencies are significantly smaller for each component (SW, LW and Net). SW heating amounts to less than 1K day^{-1} across all of the extratropics (effectively zero close to the pole which is subject to polar-night conditions) and LW cooling ranges from $\sim 0\text{K day}^{-1}$ close to 30°S to between 1 and 2K day^{-1} over much of the polar-cap region (60 - 90°N). Therefore, the Net tendency (Figure 29c) more closely resembles that of the spatial pattern in LW cooling (slight offset from SW heating) and is generally negative in sign, consistent with the weak solar input at this time of the year. The difference in SW heating between the two forecasts (Figure 29e) shows some interesting spatial structure (although more muted compared with ~ 10 hPa). Over eastern Russia, a local positive anomaly in ozone (as shown before in Figure 28a) coincides with a positive residual in SW heating (up to 0.1K day^{-1}). This association is indeed applicable elsewhere to a smaller extent, except close to the pole where sunlight is absent. Remaining regions characterised by a negative residual in SW heating (dSW), such as is evident over the eastern Pacific, US, western North Atlantic, south-western Europe and southern Russia (~ -0.1 to -0.25K day^{-1}), furthermore correspond well with the distribution of negative ozone anomalies. The LW cooling difference (dLW) (Figure 29d) tends to be inverse in sign to that of dSW over high-latitudes, as well as eastern Russia, where the anomaly in ozone is strongly positive. This is decreasingly the case at lower latitudes, with the dLW largely of similar sign and magnitude to the dSW near 30°N , meaning the radiative impact has an additive impact on temperature. The Net tendency difference (dNet) spatial pattern (Figure 29f) overall appears similar to the dSW field, which is modulated by influence from the dLW as described above. In absolute terms, the radiative impact of the HLO scheme is

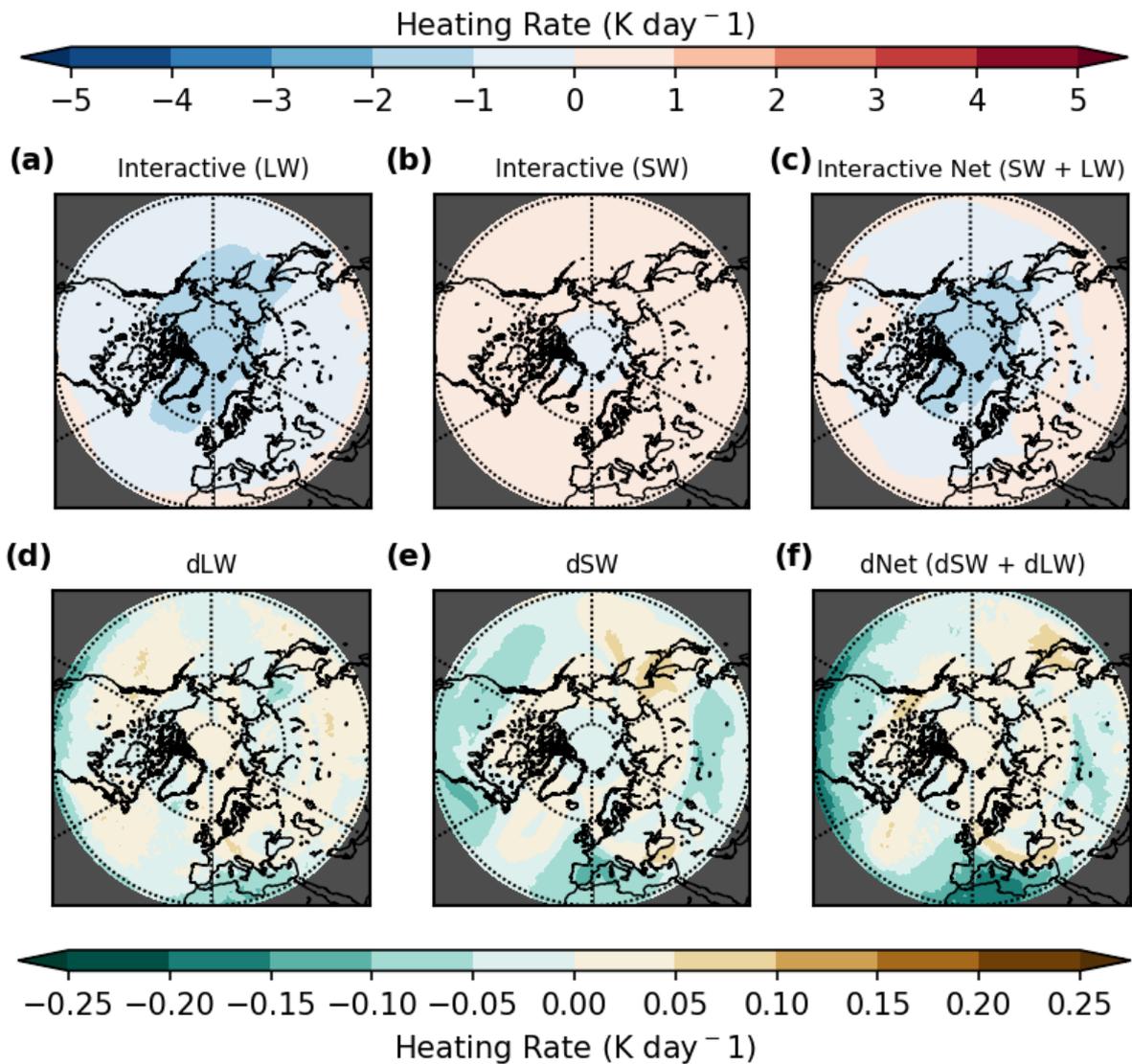


Figure 29: Same as Figure 27 but for the 70 hPa level.

computed to be smaller than at 10 hPa (at least over mid-latitudes) but can be inferred to be relatively larger at 70 hPa as the Net radiative heating tendency is significantly smaller than at 10 hPa.

5.2. Correlations and RMSE verification

To assess the radiative importance of ozone on temperature as function of pressure (altitude) and lead time throughout the full forecast (60 days), a series of correlations are computed over the extratropics for each hemisphere for this specific event. Variables correlated (calculated for the ensemble-mean) include the temperature difference (interactive minus control); the interactive anomaly in ozone; the difference in radiative heating tendencies (dSW, dLW and dNet); as well as both interactive ozone and temperature anomalies with respect to analysis. As an addition to this, the RMSE skill metric is

computed to help discern whether the HLO scheme leads to an overall improvement (reduction) in temperature forecast error.

First in Figure 30, the evolution in the Pearson correlation coefficient (r) between the forecast difference in temperature (dT) and the residuals in radiative heating tendencies (dSW , dLW and $dNet$), is shown for the nearest model levels to 10, 30, 50, 70 and 100 hPa, calculated using all model gridpoints poleward of $20^\circ N$ and $20^\circ S$ to represent the northern hemisphere (panels a-c) and the southern hemisphere (panels d-f) extratropics respectively. For both hemispheres, the dT is positively correlated with the dSW heating, with a few minor exceptions. In the northern hemisphere, the correlation is strongest generally in the lower stratosphere (~ 70 hPa), where $r = \sim 0.6$ except briefly around $\sim 3-4$ weeks into the forecast. The weakest correlation is shown at 10 hPa, which remains below 0.4 throughout and as low as ~ 0 during the first two weeks and is probably a result of strong adiabatic descent during this time (dynamically induced warming). Overall, the correlation tends to increase slightly at every level as the forecast evolves, and this likely reflects the growing importance of SW heating as winter

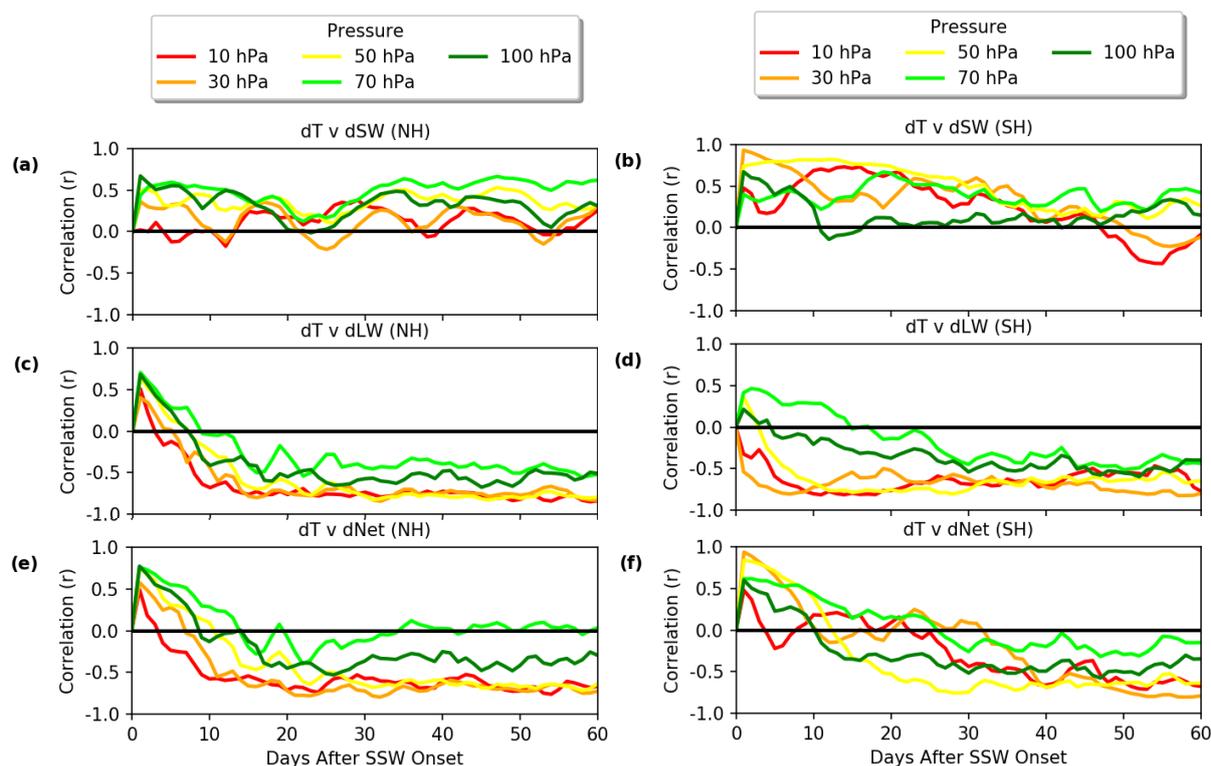


Figure 30: Temporal evolution of the correlation (r) between the ensemble-mean temperature difference with respect to the difference in (a,b) SW (dSW), (c,d) LW (dLW) and (e,f) Net ($dNet$) radiative heating tendencies, between the interactive and control forecast experiments, for (left) the northern hemisphere extratropics ($20-90^\circ N$) and (right) the southern hemisphere extratropics ($20-90^\circ S$). Both sets of forecasts were initialised on the central warming date of the 24th of January 2009 SSW event.

transitions to spring. The opposite is true for the southern hemisphere, which again reflects seasonal transition from summer to autumn. Here, the declining strength in the correlation is most pronounced in the mid-stratosphere (~10 hPa), as high as 0.8-0.9 in first couple of weeks but weakly negative by the forecast end. In contrast, a weak to moderate positive correlation largely persists in the lower stratosphere (70 and 100 hPa), albeit gradually declining.

The evolution in the dT versus dLW heating correlation for each level is very similar for both hemispheres. In the northern hemisphere, an initial positive correlation (up to $r \sim 0.6$) quickly becomes negative a week into the forecast, largely levelling off for the remainder of the forecast duration from two weeks lead time. The anticorrelation is strongest at both 10 and 30 hPa ($r \sim -0.8$) and weakest around 70 hPa ($r \sim -0.4$). Over the southern hemisphere, the initial correlation similarly starts positive in the lower stratosphere ($r \sim 0.3$), before reverting to a weak to moderate negative correlation at 70 and 100 hPa ($r = -0.4$ to -0.5 after 30 days), but quickly becomes strongly negative ($r \sim -0.8$) within days in the mid-stratosphere (10 and 30 hPa). The increasing negative correlation as the forecast evolves reflects the LW heating response to the SW heating driven temperature difference between the ensemble-mean of each experiment, which opposes that of the temperature signal. The dT versus dNet correlations closely resemble that of the correlations of dT versus dLW heating in the northern hemisphere, albeit weakened significantly in the lower stratosphere (70 and 100 hPa), which reflects the relatively minor importance of SW heating during late winter. The equivalent dT versus dNet correlations in the southern hemisphere conversely reflect the dominance of the SW heating component early in the forecast (moderately positive for the first 7 to 10 days), which progressively becomes negative for all levels 2 to 4 weeks into the forecast. Whilst the evolution of the dT versus dNet correlations are sensitive to seasonal transition (particularly over the southern hemisphere), the changes through time can be largely explained by the LW cooling response to existing differences in temperature between the two forecasts.

Correlations were next performed between the anomaly in ozone (as simulated by the interactive forecast) and the radiative heating tendency differences (displayed in Figure 31), again for the northern hemisphere (20-90°N) and southern hemisphere (20-90°S) extratropics. Over the northern hemisphere, a moderate to strong positive correlation is found for the projected ozone anomaly with respect to the dSW, which largely increases as the forecast evolves ($r \sim 0.6-0.9$). Despite the weak influence of the SW heating component during winter, it is obvious that higher ozone at each level is associated with increased SW heating. The result is similar over the southern hemisphere, albeit the declining solar input with time means the correlations overall weaken slightly as the forecast evolves. Between two and four weeks into the forecast however, the strong positive association between ozone anomaly and SW heating is interrupted at 10 hPa (r becomes negative) and significantly weakened at both 70 hPa,

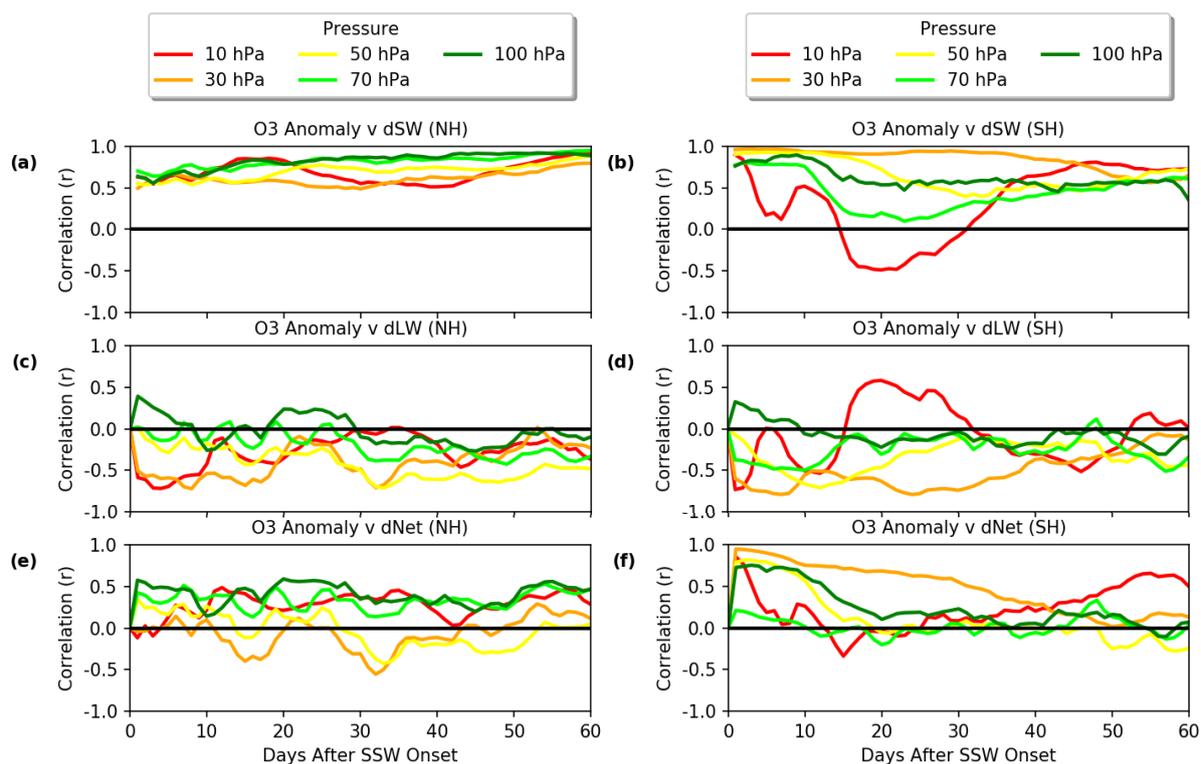


Figure 31: Temporal evolution of the correlation (r) between the ensemble-mean interactive ozone evolution with respect to the difference in (a,b) SW (dSW), (c,d) LW (dLW) and (e,f) Net (dNet) radiative heating tendencies, between the interactive and control forecast experiments, for (left) the northern hemisphere extratropics (20-90°N) and (right) the southern hemisphere extratropics (20-90°S). Both sets of forecasts were initialised on the central warming date of the 24th of January 2009 SSW event.

and to a lesser extent, 100 hPa. The reason for this is unclear but could be associated with non-local radiative effects overcompensating for local radiative control, which would be supported by a similar deviation in the overall correlation for ozone versus dLW here during this time frame.

The association between the interactive simulated ozone anomaly and the dLW tendency is generally negative in each hemisphere, with the strength of the anticorrelation highly variable between levels. As explained, an indirect cause of such anticorrelation arises from the perturbation to temperature that ozone may yield (via modulated SW heating) and thus LW cooling is attenuated to help restore radiative equilibrium. A direct mechanism that would further support such strong negative correlation would be the availability or more (less) ozone to emit, leading to stronger (weaker) cooling in accordance with an enhanced (reduced) greenhouse effect. The relative importance of both such mechanisms cannot be quantified here, but both contribute to the strong anticorrelations shown here. The correlation of the ozone anomaly with the dNet tendency mimics that of the correlation with the dLW, although the correlation is instead overall weakly positive. Therefore, the ozone anomaly in each hemisphere tends

to be more frequently positively associated with an overall warming tendency as the SW heating (direct effect) prevails slightly over the resultant LW cooling (indirect effect). This evaluation however only serves to demonstrate the relative impact of these two radiative components and does not account for the absolute conditions as in section 5.1 (determined by solar input).

In Figure 32, the correlation between the interactive ozone anomaly and the temperature difference (with respect to the control forecast) is subsequently shown for each level, calculated over the northern (20-90°N) and southern hemisphere (20-90°S) extratropics as before. The temporal evolution of these correlations highlights the variable importance of the HLO scheme, in place of a zonal-mean, monthly-mean ozone climatology as seen by the radiation scheme, in influencing the ΔT between the two experiments. As this constitutes the only difference between the two sets of initialised forecasts, it can be inferred that where the correlation drops, feedback processes (involving perturbations to stratospheric dynamics) serve to counter the radiative effects of ozone.

To accompany this set of evaluations (Figures 30 to 32), the evolution in the interactive ozone and forecast temperature (for both experiments) correlation, with respect to the ERA-5 analysis, is computed for each pressure level, together with the RMSE, again over the northern and southern hemisphere extratropics (poleward of 20° latitude), as displayed in Figures 33 and 34 respectively. Note that all variables were first calculated as anomalies with respect to the ERA-5 monthly-mean climatology used

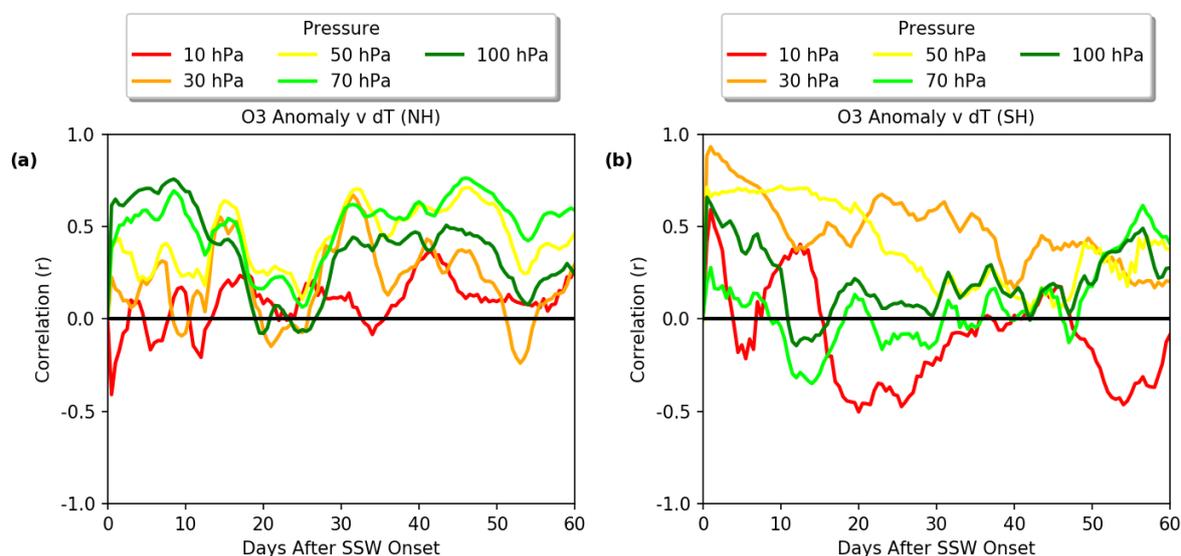


Figure 32: Temporal evolution of the correlation (r) between the ensemble-mean interactive ozone evolution with respect to the temperature difference (ΔT) between the interactive and control forecast experiments for (a) the northern hemisphere extratropics (20-90°N) and (b) the southern hemisphere extratropics (20-90°S). Both sets of forecasts were initialised on the central warming date of the 24th of January 2009 SSW event.

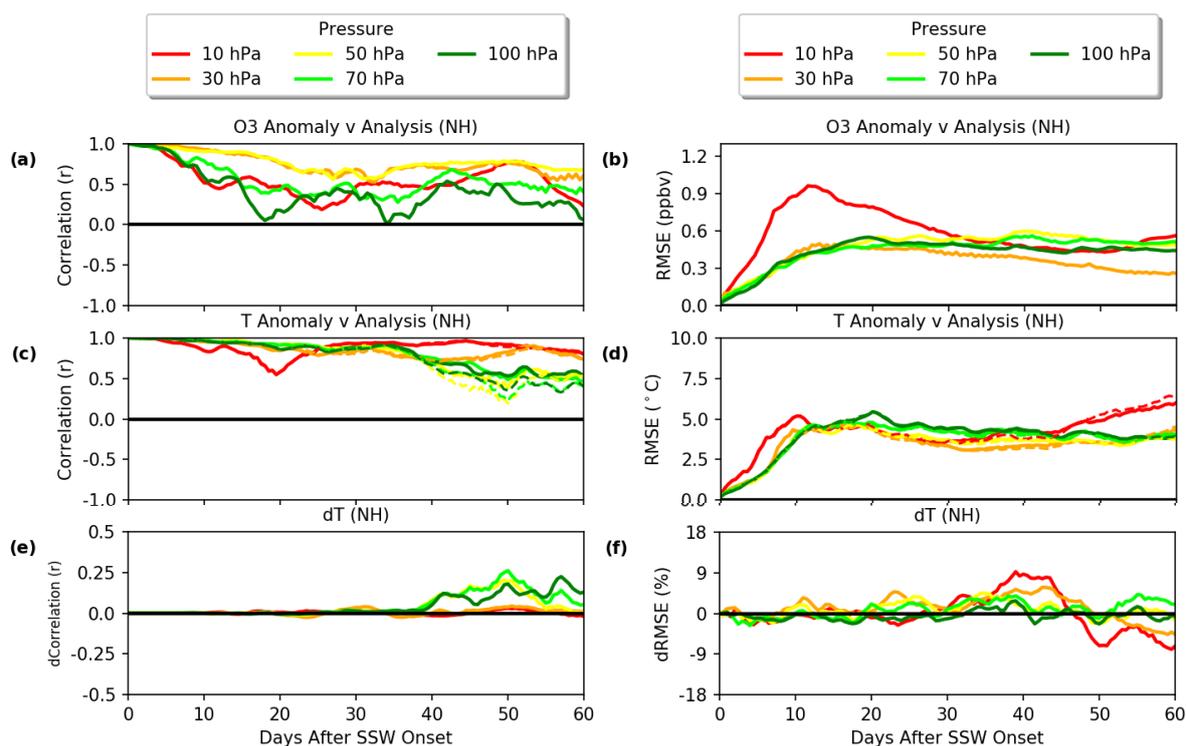


Figure 33: Northern hemisphere (20-90°N) temporal evolution of the correlation (r) between the: (a) ensemble-mean interactive ozone evolution and ERA-5 ozone analysis; (c) ensemble-mean interactive temperature anomaly and ERA-5 temperature analysis and (e) the interactive minus control ensemble-mean temperature correlation difference. The computed area-weighted RMSE is subsequently shown: (b) for ozone; (d) for temperature and (f) for the interactive minus control normalised RMSE difference (%).

(2000-2013). For the northern hemisphere (Figure 33), the correlation between the forecast and analysis anomaly in ozone steadily falls away in the first 3-4 weeks of the forecast, with generally little change thereafter. The accompanying RMSE evolution matches the broad flattening out of the correlations from 25 days onwards (~ 0.5 ppmv for levels 30, 50, 70 and 100 hPa), but begins to decrease slightly again after 40 days (as spring approaches). The RMSE at 10 hPa however reaches up to ~ 1.0 ppmv between approximately 10 and 20 days into the forecast, which likely reflects the outfall from the SSW onset (which involves large poleward and downward transport of ozone into the polar-cap region), which is most pronounced at this pressure (altitude). The RMSE however falls markedly thereafter to within the range seen for the other examined levels.

During the first 30 days of the forecast, the evolution in the correlation between the projected and analysis temperature can be related closely to that shown for ozone (in support of a positive impact from radiatively interactive HLO). As an example of this, a decrease in the correlation strength at 50, 70 and 100 hPa after 40 days is significantly reduced (by between ~ 20 and 40%) for the interactive experiment.

However, the normalised difference in RMSE (dRMSE) between the two forecasts is far less stark during this time, despite a very slight overall signal for improvement (negative residual in RMSE). This suggests that in the later stages of the forecast, radiatively interactive HLO helps to improve accuracy in the spatial distribution of temperature variability, whilst having a largely negligible impact on the magnitude of the temperature difference. This is however at odds with the interpretation made in section 3.3 for medium-range results of the HLO scheme performance following the January 2018 SSW, in which 5-day RMSE and ACC scores in forecast temperature essentially implied the opposite (improvement in mean forecast error but not pattern correlation). Although this merits further investigation, in conjunction with the residence of a long-lived lower stratospheric positive anomaly in ozone following such events (as shown in section 4.1 to 4.3), such discrepancy could be a function of lead time and the differential benefit HLO may bring over different forecast timescales. The RMSE for the interactive forecast is on average marginally smaller (~1-3%) during the first couple of weeks too, but tends to be slightly greater between 20 to 40 days. The greatest temporal variance in the temperature RMSE is shown for the 10 hPa level (up to $\pm 9\%$ after 30 days).

The same plots are next shown over the southern hemisphere extratropics (Figure 34). As for the northern hemisphere, the correlation between the anomaly in interactive forecast and analysis ozone steadily falls in the first 2-3 weeks before levelling off ($r = \sim 0.2-0.6$ after 25 days); remaining highest in the mid-stratosphere throughout much of the forecast. Due to the difference in ozone volume mixing ratio with pressure (altitude), the RMSE in projected ozone rises only very slowly at 50, 70 and 100 hPa (only 0.15-0.25 ppmv after 60 days), but much more significantly at both 10 and 30 hPa (>0.5 ppmv after 30 days). Such characteristics reflect the greater spatial variability in ozone anomalies within the lower stratosphere, albeit relatively small in absolute terms, and relatively homogeneous spatial anomaly field in the mid-stratosphere (as was demonstrated in section 5.1.1). The interactive and control forecast temperature anomaly versus analysis correlations are characterised by a steady drop in the first month of the forecast for all levels, with weak to negligible spatial predictability after 30 days lead time ($r = <0.4$). A strong diurnal signal is apparent in the correlation temporal evolution (especially towards 10 hPa). This suggests that some aspect of the diurnal cycle, perhaps related to the photochemical evolution of ozone, is not well represented by either the interactive or control experiment and should be separately investigated. The correlation difference evolution (Figure 34e) shows an overall slight degradation in the performance of the interactive forecast after 10 days, in terms of the spatial placement of temperature variations, with tentative indication of neutral impact from radiatively interactive HLO at later times. The temperature RMSE evolution (Figure 34d) rapidly increases in the first 10 days, particularly in the lower stratosphere ($>3^\circ\text{C}$ after 10 days at 100 hPa), and more steadily thereafter (always smallest at 10 hPa). In conjunction with the correlation difference between the ensemble-mean of the two experiments, the RMSE temperature difference (Figure 34f) also suggests an overall

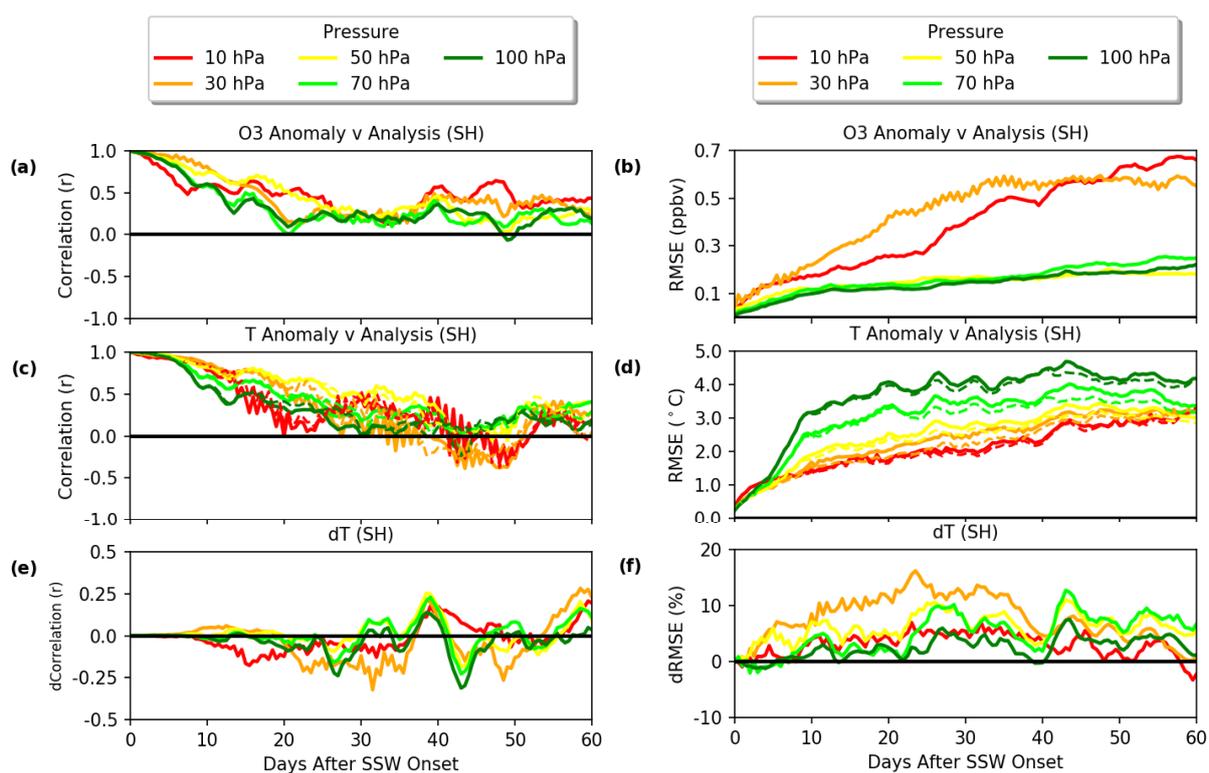


Figure 34: Southern hemisphere (20-90°S) temporal evolution of the correlation (r) between the: (a) ensemble-mean interactive ozone evolution and ERA-5 ozone analysis; (c) ensemble-mean interactive temperature anomaly and ERA-5 temperature analysis and (e) the interactive minus control ensemble-mean temperature correlation difference. The computed area-weighted RMSE is subsequently shown: (b) for ozone; (d) for temperature and (f) for the interactive minus control normalised RMSE difference (%).

deterioration from implementing HLO, interactive with radiation, in absolute terms (by up to 30%). The robustness of this finding ought to be tested following other such events, as well as non-SSW periods for comparison purposes, but seems to be primarily associated with unfavourable changes to stratospheric mean temperature biases due to HLO (based on earlier results of this report).

6. Synthesis, recommendations and future work

In this section, a series of recommendations concerning both the near-term operational implementation of the radiatively interactive HLO scheme, as well as ongoing subsequent development, in the IFS are provided. Particular emphasis is given to different forecast timescales (e.g. medium range through to sub-seasonal) covered by separate NWP systems at ECMWF, based on the findings of this investigation. Where further verification of the HLO scheme and its impact is however believed to be warranted, suggested actions are provided to guide future work necessary to inform usage of radiatively interactive HLO, prior to operational implementation.

In section 1.4, some key research questions were first outlined to motivate the work included in this report. These are quoted below in turn, followed by a detailed response that attempts to answer these questions based on the findings here:

- 1) *“How does radiatively active prognostic ozone, using the HLO scheme, influence forecast skill in the IFS on medium-range (up to 15 day) and sub-seasonal (up to 60 day) timescales? How does the impact compare between each hemisphere (e.g. summer versus winter)?”*

The evidence from this report is that radiatively interactive HLO yields an overall positive impact on forecast skill in the medium range, particularly focussed in the summer hemisphere. Averaged over a recent 11.5-month period between April and September, it was shown in section 3.1 that HLO improved the RMSE in temperature by up to 4% and the ACC by as much as ~10% between 10 and 100 hPa over the northern hemisphere extratropics for a forecast lead time of three days (T+72). In contrast, the bias in mean temperature and wind (via the thermal-wind balance relationship) is found to degrade in the tropical upper stratosphere (~1-10 hPa) region primarily due to radiatively interactive HLO; the RMSE degrades by ~5-10% for temperature and ~1-2% for wind over this same period for a 3-day forecast. The association of skill change with enhanced solar input is expected due to the greater importance of SW heating, which constitutes the direct radiative response to spatial variations in ozone abundance. The signal for improvement is clearly complicated by interaction with mean biases in the IFS. The indication is that whilst some features such as a LMS cold bias in the IFS during winter (centred near 200 hPa) are reduced, other biases are negatively impacted by radiatively interactive HLO (e.g. near-stratopause warm bias). Isolation of the variability aspect of skill change using metrics such as SDE and ACC, which are insensitive to changes in mean bias, however, presents a clearer signal for overall improvement in temperature and wind representation due to radiatively interactive HLO. The results are consistent for this forecast range when examining sub-seasonal ensemble experiment data, initialised on the central warming dates of six major midwinter, PJO-class SSW events (in which the lower stratosphere following such events is significantly perturbed relative to climatology). After ~10-

15 days however, the impact due to HLO, interactive with radiation, becomes more complicated to understand and warrants closer examination.

- 2) *“How do projected spatial anomalies in ozone affect forecast temperature? Through thermal-wind balance, what impact does this subsequently have on spatial wind patterns?”*

As was first shown in section 3.2 for an individual case during the quiescent state of the summertime southern hemisphere stratosphere at a 7-day forecast lead time, spatial anomalies in ozone have a clear correspondence with induced temperature changes, as a result of making HLO radiatively interactive. This example highlights the importance of the local, direct SW heating response. Via the thermal-wind balance relationship, wind velocity and direction are modified by alterations to horizontal temperature gradients in accordance with expectations. Later investigation in conjunction with radiative heating tendencies (undertaken in section 5.1) for another individual case (forecast for 7 days after the 24th of January 2009 SSW), largely supports the findings over the southern hemisphere (although the comparison is limited as they cover two distinct, individual snapshots). The SW heating response is overall dominant in the summer hemisphere, despite a partial offset due to LW cooling, in response to induced deviations in temperature. The direct (thermal emission dependence due to ozone abundance) and indirect (thermal emission dependence due to ambient temperature) LW cooling response mechanisms highlighted in section 5 lead to such compensation, which appears to equilibrate after approximately ~2-3 weeks into the forecast. The relative importance of each mechanism ought to be investigated further to understand the limitations of the benefit HLO, interactive with radiation, may yield on sub-seasonal time frames.

In contrast, in the perturbed northern hemisphere stratosphere following the most prolonged, long-lived SSW on record (Manney et al., 2009a), the benefit of HLO is difficult to ascertain due to the manifestation of small-scale structures in ozone which are inherently difficult to predict at even relatively short lead times (e.g. 7 days or T+168). Nonetheless, SW heating appears to have some importance over mid-latitude regions at this time, with evidence of some improvement due to HLO.

- 3) *“Centred on northern hemisphere midwinter sudden stratospheric warming (SSW) events, how does prognostic HLO, interactive with radiation, impact forecast skill following such events? Is the dynamical coupling of stratospheric skill into the troposphere positively impacted by such implementation?”*

The indication from evaluation of HLO performance in both the medium and sub-seasonal range is that HLO reduces temperature forecast bias in the middle to upper stratosphere leading up to and during an SSW onset. This is captured by skill metrics sensitive to both change in the mean bias and variability (e.g. RMSE), and those which score only the change in pattern correlation skill (e.g. ACC). For instance, it was shown in section 3.3 that the RMSE in temperature was improved by ~3-6% and the ACC by

greater than 2% during this time between ~1 and 10 hPa, around the timing of the February 2018 SSW. An overall signal for slight improvement propagates down into the lower stratosphere following the onset, within approximately 1-2 weeks, and resides in the lower stratosphere on sub-seasonal forecast timescales (demonstrated out to 60 days in experiments performed here). Although not studied here in detail, it is during this period when an improvement in the predictability of stratosphere-troposphere coupling mechanisms may be anticipated due to radiatively interactive HLO. The signal at later forecast lead times (e.g. 30-60 days) is less clear but a beneficial, radiative impact of a positive lower stratospheric ozone anomaly (~70-100 hPa) that persists, according to that simulated using the HLO scheme and both reanalysis data from ERA-5 and CAMS (which is a consistent feature across all events), is implied. Conversely, a tendency for a slight overall skill degradation was found for the 20-40 day forecast range, which reverts to an overall positive impact 40-60 days following the event. The robustness of this tendency is limited, however, by the specific number of events examined but could reflect the changes associated with the event aftermath. At longer forecast lead times, any difference in signal between summer and winter hemispheres is not as obvious, unlike that for the medium range.

- 4) *“What are the radiative mechanisms by which the typical polar-cap evolution of ozone following PJO-type SSWs affects the forecast response? How does this impact translate to skill change spatially (both vertically and horizontally)?”*

The in-depth analysis of the 2009 SSW evolution, in conjunction with radiative heating tendency output, demonstrated the difficulty of ascertaining the impact of radiatively interactive HLO following the event. A particular issue is the large, dynamically-induced changes in temperature ($>10^{\circ}\text{C}$), corresponding to enhanced poleward and downward descent, which compares with cumulative heating changes due to ozone of typically less than 1°C . Whilst SW heating associated with streamers of ozone-rich air over mid-latitudes (advected from lower latitudes) was shown to have a beneficial impact on temperature prediction here, the impact was not clear at high-latitudes where the anomalous build-up of ozone is most prominent. This is likely a result of the relative absence of SW heating and dominance of any indirect LW cooling signal, which is complicated by radiative relaxation following a period of sustained adiabatic descent around the onset date, which results in such abrupt warming. A major limitation of this investigation is the phase shift in ozone anomaly structure at even relatively short lead times. For instance, at a forecast time of 7 days (T+168), projected mid-stratospheric ozone structure following the 2009 SSW was largely out of phase with that according to the ERA-5 analysis. However, this apparent disagreement likely reflects the extreme conditions following the event and is unlikely to be representative of the performance during more typical conditions.

Examination of correlations between ozone, forecast temperature and radiative heating tendency differences furthermore highlights another possible reason for the lack of improvement signal due to

radiatively interactive HLO on sub-seasonal timescales. A strong anticorrelation ($r < -0.6$) between the difference in the daily-mean LW heating rate between the two sets of forecasts (interactive minus control), with respect to the forecast difference in temperature, establishes some 10-15 days after the event onset (forecast initialisation date). Although this complicates the association of the temperature response due to SW heating, the evaluations presented here (in section 5.2) provide a useful diagnosis of why this occurs, and the time frame at which an equilibrium is reached between the ozone/SW heating driven temperature change and the counteractive LW cooling negative feedback. Such understanding provides scope to extend the utility of HLO, interactive with radiation, on forecast timescales beyond the medium range. Despite such existing limitations, the radiative heating tendency output nonetheless supports an enhancement in temperature skill at longer lead times (e.g. 40-60 days) due to a residual positive anomaly in lower stratospheric ozone. This skill gain may emerge as spatial variability in ozone reduces in accordance with SPV recovery and as the solar input increases with the seasonal transition from winter to spring.

To address these questions further, a series of suggested actions are provided below:

- The HLO scheme could be refined to further enhance the realism of the ozone evolution following SSW events in particular. 3D forecast fields in ozone may be limited by both errors in the initial ozone analysis, which is likely to be a strongly influential factor in the short range, as well as inadequacies in the linear ozone scheme performance, which becomes an increasingly important limitation at longer forecast lead times. This first requires more detailed quantitative study, with the former aspect testable through validation against in situ (e.g. ozonesonde) observations directly, or indirectly with reference to independent satellite observations (not assimilated in CAMS), and the latter by comparing simulated output against full chemistry scheme output. It should be noted that in this report, evaluations comparing the performance of radiatively prognostic ozone derived using the HLO scheme with other schemes (e.g. CD and BMS) have not been performed, and could thus also be facilitated with the aid of such observational sources. For all 6 PJO-type SSW events studied (2000-2020), a tendency to underestimate both the magnitude and duration of positive ozone anomaly features was found (using the HLO scheme). An aspect influencing this could be a tendency within IFS to promote faster recovery of the SPV than observed, perhaps as a consequence of the inherent cold bias in the polar mid- to lower-stratosphere during winter.
- The radiative impact of a persistent positive lower stratospheric ozone anomaly (~70-100 hPa out to 60 days) following PJO-type SSWs over the Arctic (60-90°N) may not be captured cleanly in these evaluations (sub-seasonal experiments only initialised on the central warming date for each event). The largest impact would be anticipated as the mid- to upper-stratosphere recovers (SPV begins to reform), whilst the lower stratosphere remains heavily perturbed (radiative recovery timescales of 2-3 months). Therefore, further investigation should focus on later initialisations to

see whether a positive impact can be robustly ascertained. The in-depth examination of the January 2009 SSW, with reference to daily-mean radiative heating rate output, demonstrated the limited scope of the HLO scheme, interactive with radiation, to yield skill improvement beyond ~2-3 weeks lead time. This is principally related to the counteractive LW cooling response to existing anomalies in ozone and temperature (negative feedback), yet the results presented might enable this limitation to be exploited. Other factors that could influence the emergence of skill would include the shift from winter to spring (sun and thus SW heating become more influential). It has also already been seen that HLO has a clear positive impact on forecast skill more generally (over multi-month periods and particularly when sunlight is more abundant) in the medium range (out to 15 days), particularly between ~10 and 100 hPa. Even during the summertime southern hemisphere for these event times, the impact is not however clear in the extended range (15-60 days). So, this likely contributes to the lack of signal found so far for these extreme events.

- A larger suite of SSW events require investigation to quantify and understand the skill improvement HLO interactive with radiation can yield. The small sample of six events covered in this report is insufficient to characterise the mean response, as each event shows a different signal. Therefore, the composite approach shown in section 4.1 may still not be truly representative of a typical response to a PJO-type SSW, due to the conflation of signals between events. A particular aspect that has not been addressed sufficiently in this report is the impact HLO may have on tropospheric predictability and improved representation of stratosphere-troposphere coupling. Oehrlein et al. (2020) for instance found that inclusion of interactive ozone chemistry in the WACCM model is pertinent to adequately reproducing the characteristics of such events, in terms of both the mean state and variability, and in explaining more prolonged surface impacts as a possible result of feedbacks between ozone and dynamics during such events when advection of ozone-rich air into polar latitudes occurs. It would be informative to apply similar evaluations using IFS forecast data, comparing the results for a radiatively interactive HLO and control experiment.
- Longer, extended periods of verification in HLO performance, particularly in the sub-seasonal forecasting range, are warranted during non-SSW periods. Although the computational costs of executing ensemble experiments limit the number of dates forecasts may be initialised on, it would be helpful to run experiments on key dates during different seasons (across multiple years) to determine the seasonal dependence of HLO impact on forecast skill. Nevertheless, other events of interest may be targeted such as final warmings, which are known to subsequently impact the troposphere (e.g. Black et al., 2006), as well as transitions in the phase of the QBO, which constitutes an important source of global-scale predictability on seasonal timescales (Ebdon, 1975; Folland et al., 2012).

7. Conclusions

The performance of the novel HLO scheme, interactive with radiation within the IFS, is evaluated in this report for a variety of individual and aggregate cases, spanning the period 2000-2020 over both medium-range (0-15 day) and sub-seasonal (0-60 day) timescales. Particular focus was devoted to understanding the impact of HLO following the evolution of northern hemisphere PJO-type midwinter SSWs, which are characterised by a profound impact on the dynamical state of the lower stratosphere. Such events are a renowned source of predictability on sub-seasonal timescales, including tropospheric weather regimes, and therefore implementations that may yield additional skill enhancement following these events are of strong interest to operational forecast centres such as ECMWF. The results outlined in this report provide supporting evidence that HLO interactive with radiation would yield a significant positive overall impact if implemented in the medium-range forecasting system. It is therefore recommended here that radiatively interactive HLO should be made operational in the next major cycle of the IFS (Cycle 48r1). The caveats to any such implementation mostly concern interaction with mean biases present in IFS, which are shown to be unfavourable in some cases, leading to local skill degradation (e.g. the tropical mid-stratosphere). However, these will be mitigated by combining them with other changes that affect the mean stratospheric temperature, namely a modification of the solar spectrum to reduce incoming ultraviolet radiation by 7-8% (Hogan et al. 2017), and the introduction of a Semi-Lagrangian Vertical Filter to suppress grid-scale temperature fluctuations in the vertical (Polichtchouk et al. 2021).

Several issues are highlighted which may explain the difficulty in attaining skill improvement beyond the medium range (>15 days), particularly following the extreme case of SSWs. A series of recommendations are outlined in section 6 to remedy and circumvent some of these issues, as well as the suggestion that the HLO scheme could be better refined to match the evolution in ozone according to both analyses (e.g. CAMS and ERA-5) and observations (e.g. ozonesondes). Close investigation of the skill changes due to radiatively interactive HLO, using four different established verification metrics, imply a more coherent signal for an improvement in the variability aspect due to HLO, relative to an enhancement in skill in terms of mean forecast bias. Although the latter is in evidence in some cases (e.g. the LMS wintertime cold bias centred near 200 hPa), other regions of the stratosphere are subject to significant degradation in mean bias. Following northern hemisphere midwinter PJO-type SSWs, this is principally a concern between 20 and 40 days into the forecast, with some benefit yielded in the first 20 days of the forecast (when the anomaly in ozone is largest and extends vertically throughout much of the stratosphere). Another coherent feature is a local skill improvement in the polar lower stratosphere (~70-100 hPa) between 40 and 60 days after an event, in terms of both variability and mean bias, in connection with a residual positive anomaly in ozone that persists and decays only

slowly (as the SPV recovers at higher altitude). This feature could impact the tropospheric response to such events during the time and warrants further investigation, as recommended in section 6. Further verification of the performance of radiatively interactive HLO on sub-seasonal timescales is merited, building on the findings of this report which highlight inherent limitations at longer lead times, that encompasses both extended non-SSW periods, as well as other events of interest such as final warmings and QBO transitions.

Acknowledgments

This project was facilitated through six months funding generously provided by the Copernicus Atmospheric Monitoring Service (CAMS) to support RSW to carry out this study. ECMWF implements the Copernicus Atmospheric Monitoring Service on behalf of the European Union. RSW would like to extend his gratitude to support provided by ECMWF, including co-author contribution from RH, IP, TS, JF and MIH (University of Reading).

References

- Abdi, H. (2007). Bonferroni and Šidák corrections for multiple comparisons. *Encyclopedia of measurement and statistics*, **3**, 103-107.
- Andersson, E. (2015). *Appendix A*. Retrieved from: http://www.atmos.albany.edu/daes/atmclasses/atm401/spring_2016/ppts_pdfs/ECMWF_ACC_definition.pdf.
- Andrews, D. G., Holton, J. R. and Leovy, C. B. (1987). *Middle atmosphere dynamics* (No. 40). Academic press.
- Arblaster, J. M. and Meehl, G. A. (2006). Contributions of external forcings to southern annular mode trends. *Journal of climate*, **19**(12), 2896-2905.
- Black, R. X., McDaniel, B. A. and Robinson, W. A. (2006). Stratosphere–troposphere coupling during spring onset. *Journal of climate*, **19**(19), 4891-4901.
- Butler, A. H., Sjoberg, J. P., Seidel, D. J. and Rosenlof, K. H. (2017). A sudden stratospheric warming compendium. *Earth System Science Data*, **9**(1), 63-76.
- Byrne, N. J., Shepherd, T. G. and Polichtchouk, I. (2019). Subseasonal-to-seasonal predictability of the Southern Hemisphere eddy-driven jet during austral spring and early summer. *Journal of Geophysical Research: Atmospheres*, **124**(13), 6841-6855.
- Cagnazzo, C., Manzini, E., Fogli, P. G., Vichi, M. and Davini, P. (2013). Role of stratospheric dynamics in the ozone–carbon connection in the Southern Hemisphere. *Climate dynamics*, **41**(11-12), 3039-3054.
- de la Cámara, A., Abalos, M. and Hitchcock, P. (2018a). Changes in stratospheric transport and mixing during sudden stratospheric warmings. *Journal of Geophysical Research: Atmospheres*, **123**, 3356-337.
- de la Cámara, A., Abalos, M., Hitchcock, P., Calvo, N., and Garcia, R. R. (2018b). Response of Arctic ozone to sudden stratospheric warmings. *Atmospheric Chemistry and Physics*, **18**(22), 16499–16513.
- Cariolle, D. and Déqué, M. (1986). Southern hemisphere medium-scale waves and total ozone disturbances in a spectral general circulation model. *Journal of Geophysical Research: Atmospheres*, **91**(D10), 10825-10846.
- Cariolle, D. and Teyssedre, H. (2007). A revised linear ozone photochemistry parametrization for use in transport and general circulation models: multi-annual simulations. *Atmospheric Chemistry and Physics*, **7**(9), 2183-2196.

- Charlton, A. J., Polvani, L. M., Perlwitz, J., Sassi, F., Manzini, E., Shibata, K. et al. (2007). A new look at stratospheric sudden warmings. Part II: Evaluation of numerical model simulations. *Journal of Climate*, **20**, 470-488.
- Checa-Garcia, R., Hegglin, M. I., Kinnison, D., Plummer, D. A. and Shine, K. P. (2018). Historical tropospheric and stratospheric ozone radiative forcing using the CMIP6 database. *Geophysical Research Letters*, **45**(7), 3264-3273.
- Cheung, J. C. H., Haigh, J. D. and Jackson, D. R. (2014). Impact of EOS MLS ozone data on medium-extended range ensemble weather forecasts. *Journal of Geophysical Research: Atmospheres*, **119**(15), 9253-9266.
- Chipperfield, M. P. (2006). New version of the TOMCAT/SLIMCAT offline chemical transport model, *Quarterly Journal of the Royal Meteorological Society*, **132**, 1179–1203.
- Coopmann, O., Guidard, V., Fourrié, N. and Josse, B. (2020). Use of variable ozone in a radiative transfer model for the global Météo-France 4D-Var system. *Quarterly Journal of the Royal Meteorological Society*, **146**(733), 3729-3746.
- Copernicus (2019). CAMS Global Forecasting System: Cycle 47r1. Retrieved from: <https://atmosphere.copernicus.eu/cycle-47r1>.
- De Grandpré, J., Ménard, R., Rochon, Y. J., Charette, C., Chabrilat, S. and Robichaud, A. (2009). Radiative impact of ozone on temperature predictability in a coupled chemistry–dynamics data assimilation system. *Monthly Weather Review*, **137**(2), 679-692.
- Derber, J. C. and Wu, W. S. (1998). The use of TOVS cloud-cleared radiances in the NCEP SSI analysis system. *Monthly Weather Review*, **126**(8), 2287-2299.
- Dietmüller, S., Ponater, M. and Sausen, R. (2014). Interactive ozone induces a negative feedback in CO₂-driven climate change simulations. *Journal of Geophysical Research: Atmospheres*, **119**(4), 1796-1805.
- Domeisen, D. I. V., Sun, L. and Chen, G. (2013). The role of synoptic eddies in the tropospheric response to stratospheric variability. *Geophysical Research Letters*, **40**(18), 4933–4937.
- Ebdon, R. A. (1975). The quasi-biennial oscillation and its association with tropospheric circulation patterns. *Meteorological Magazine*, **104**, 282–297.
- Fahey, D., Newman, P. A., Pyle, J. A., Safari, B., Chipperfield, M. P., Karoly, D. et al. (2018). Scientific Assessment of Ozone Depletion: 2018, Global Ozone Research and Monitoring Project-Report No. 58.

- Flemming, J., Huijnen, V., Arteta, J., Bechtold, P., Beljaars, A., Blechschmidt, A. M. et al. Tsikerdekis, A. (2015). Tropospheric chemistry in the Integrated Forecasting System of ECMWF. *Geoscientific model development*, **8**(4), 975-1003.
- Folland, C. K., Scaife, A. A., Lindesay, J. and Stephenson, D. B. (2012). How potentially predictable is northern European winter climate a season ahead? *International Journal of Climatology*, **32**(6), 801-818.
- Geer, A. J. (2016). Significance of changes in medium-range forecast scores. *Tellus A: Dynamic Meteorology and Oceanography*, **68**(1), 30229.
- Gilford, D. M., Solomon, S. and Portmann, R. W. (2016). Radiative impacts of the 2011 abrupt drops in water vapor and ozone in the tropical tropopause layer. *Journal of Climate*, **29**(2), 595-612.
- Grooß, J. U., Konopka, P. and Müller, R. (2005). Ozone chemistry during the 2002 Antarctic vortex split. *Journal of the atmospheric sciences*, **62**(3), 860-870.
- Haase, S. and Matthes, K. (2019). The importance of interactive chemistry for stratosphere–491 troposphere coupling. *Atmospheric Chemistry Physics*. **19**, 3417-3432.
- Hansen, J., Sato, M. K. I., Ruedy, R., Nazarenko, L., Lacis, A., Schmidt, G. A. et al. (2005). Efficacy of climate forcings. *Journal of Geophysical Research: Atmospheres*, **110**(D18).
- Hersbach, H. (2000). Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, **15**(5), 559-570.
- Hitchcock, P., Shepherd, T. G. and Manney, G. L. (2013). Statistical characterization of Arctic polar-night jet oscillation events. *J. Climate*, **26**, 2096-2116.
- Hogan, R. J., Ahlgrinnm G. B., Beljaars, A., Berrisford, P., Bozzo, S., Giuseppe, F. D. et al. (2017). Radiation in numerical weather prediction. *ECMWF Technical Memoranda*, **816**, doi:10.21957/2bd5dkj8x.
- Huijnen, V., Williams, J., Weele, M. V., Noije, T. V., Krol, M., Dentener, F. et al. (2010). The global chemistry transport model TM5: description and evaluation of the tropospheric chemistry version 3.0. *Geoscientific Model Development*, **3**(2), 445-473.
- Inness, A., Ades, M., Agusti-Panareda, A., Barré, J., Benedictow, A., Blechschmidt, A. M. V. et al. (2019). The CAMS reanalysis of atmospheric composition. *Atmospheric Chemistry and Physics*, **19**(6), 3515-3556.
- Jackson, D. and Saunders, R. (2002). Ozone data assimilation system: Preliminary system, United Kingdom Meteorological Office Forecasting Research Technical Report No. 394, 18pp.

- Jeong, G. R., Monge-Sanz, B. M., Lee, E. H. and Ziemke, J. R. (2016). Simulation of stratospheric ozone in global forecast model using linear photochemistry parametrization. *Asia-Pacific Journal of Atmospheric Sciences*, **52**(5), 479-494.
- John, V. O. and Buehler, S. A. (2004). The impact of ozone lines on AMSU-B radiances. *Geophysical research letters*, **31**(21).
- Kidston, J., Scaife, A. A., Hardiman, S. C., Mitchell, D. M., Butchart, N., Baldwin, M. P. et al. (2015). Stratospheric influence on tropospheric jet streams, storm tracks and surface weather. *Nature Geoscience*, **8**(6), 433–440.
- Konopka, P., Grooß, J. U., Bausch, S., Müller, R., McKenna, D. S., Morgenstern, O. et al. (2003). Dynamics and chemistry of vortex remnants in late Arctic spring 1997 and 2000: Simulations with the Chemical Lagrangian Model of the Stratosphere (CLaMS). *Atmospheric chemistry and physics*, **3**(3), 839-849.
- Krishnamurti, T. N., Rajendran, K., Vijaya Kumar, T. S. V., Lord, S., Toth, Z., Zou, X. et al. (2003). Improved skill for the anomaly correlation of geopotential heights at 500 hPa. *Monthly Weather Review*, **131**(6), 1082-1102.
- Kuroda, Y. and Kodera, K. (2004). Role of the Polar-night Jet Oscillation on the formation of the Arctic Oscillation in the Northern Hemisphere winter. *Journal of Geophysical Research: Atmospheres*, **109**(D11).
- Lacis, A. A., Wuebbles, D. J. and Logan, J. A. (1990). Radiative forcing of climate by changes in the vertical distribution of ozone. *Journal of Geophysical Research*, **95**(D7), 9971–9981.
- 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, University of Reading, UK.
- Lu, H., Hitchman, M. H., Gray, L. J., Anstey, J. A. and Osprey, S. M. (2020). On the role of Rossby wave breaking in the quasi-biennial modulation of the stratospheric polar vortex during boreal winter. *Quarterly Journal of the Royal Meteorological Society*, **146**(729), 1939-1959.
- Manney, G. L., Schwartz, M. J., Krüger, K., Santee, M. L., Pawson S., Lee, J. N. et al. (2009a). Aura Microwave Limb Sounder observations of dynamics and transport during the record-breaking 2009 Arctic stratospheric major warming. *Geophysical Research Letters*, **36**, L12815.
- Manney, G. L., Harwood, R. S., Mackenzie, I. A., Minschwaner, K., Allen, D. R., Santee, M. L. et al. (2009b). Satellite observations and modeling of transport in the upper troposphere through the lower

mesosphere during the 2006 major stratospheric sudden warming. *Atmospheric Chemistry Physics*, **9**, 4775–4795.

Marsh, D. R., Lamarque, J. F., Conley, A. J. and Polvani, L. M. (2016). Stratospheric ozone chemistry feedbacks are not critical for the determination of climate sensitivity in CESM1 (WACCM). *Geophysical Research Letters*, **43**(8), 3928-3934.

Matheson, J. E. and Winkler, R. L. (1976). Scoring rules for continuous probability distributions. *Management Science*, **22**(10), 1087-1096.

Mathison, C., Jackson, D. R. and Keil, M. (2007). Methods of improving the representation of ozone in the Met Office model. *Met Office Forecasting Research Technical Report*, **502**, 27.

McCormack, J. P., Eckermann, S. D., Siskind, D. E. and McGee, T. J. (2006). CHEM2D-OPP: A new linearized gas-phase ozone photochemistry parametrization for high-altitude NWP and climate models. *Atmospheric Chemistry and Physics*, **6**(12), 4943-4972.

McIntyre, M. E. and Palmer, T. N. (1983). Breaking planetary waves in the stratosphere. *Nature*, **305**(5935), 593-600.

Monge-Sanz, B. M., Chipperfield, M. P., Cariolle, D. and Feng, W. (2011). Results from a new linear O₃ scheme with embedded heterogeneous chemistry compared with the parent full-chemistry 3-D CTM. *Atmospheric Chemistry and Physics*, **11**(3), 1227-1242.

Monge-Sanz, B. M., Bozzo, A., Byrne, N., Chipperfield, M. P., Diamantakis, M., Flemming, J. et al. (2021). A stratospheric prognostic ozone for seamless Earth System Models: performance, impacts and future. *Atmospheric Chemistry and Physics Discussions*, 1-39.

Myhre, G., Shindell, D., Bréon, F.-M., Collins, W., Fuglestedt, J., and Huang, J. (2013). Anthropogenic and Natural Radiative Forcing, 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*, edited by: Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 659– 740, <https://doi.org/10.1017/CBO9781107415324.018>.

Nowack, P. J., Abraham, N. L., Maycock, A. C., Braesicke, P., Gregory, J. M., Joshi, M. M. et al. (2015). A large ozone-circulation feedback and its implications for global warming assessments. *Nature climate change*, **5**(1), 41-45.

Nowack, P. J., Braesicke, P., Luke Abraham, N. and Pyle, J. A. (2017). On the role of ozone feedback in the ENSO amplitude response under global warming. *Geophysical research letters*, **44**(8), 3858-3866.

- Oehrlein, J., Chiodo, G. and Polvani, L. M. (2020). The effect of interactive ozone chemistry on weak and strong stratospheric polar vortex events. *Atmospheric Chemistry and Physics*, **20**(17), 10531-10544.
- Peuch, A., Thépaut, J. N. and Pailleux, J. (2000). Dynamical impact of total-ozone observations in a four-dimensional variational assimilation. *Quarterly Journal of the Royal Meteorological Society*, **126**(566), 1641-1659.
- Polichtchouk, I., Shepherd, T. G., Hogan, R. J. and Bechtold, P. (2018). Sensitivity of the Brewer–Dobson circulation and polar vortex variability to parametrized nonorographic gravity wave drag in a high-resolution atmospheric model. *Journal of the Atmospheric Sciences*, **75**(5), 1525-1543.
- Polichtchouk, I., Bechtold, P., Bonavita, M., Forbes, R., Healy, S., Hogan, R. et al. (2021). Stratospheric modelling and assimilation. *ECMWF Technical Memoranda*, **877**, doi:10.21957/25hegfoq.
- Randel, W. J. and Wu, F. (2010). The polar summer tropopause inversion layer. *Journal of the Atmospheric Sciences*, **67**(8), 2572-2581.
- Rao, J., Garfinkel, C. I. and White, I. P. (2020). Predicting the downward and surface influence of the February 2018 and January 2019 sudden stratospheric warming events in subseasonal to seasonal (S2S) models. *Journal of Geophysical Research: Atmospheres*, **125**(2).
- Reichler, T., Kim, J., Manzini, E. and Kröger, J. (2012). A stratospheric connection to Atlantic climate variability. *Nature Geoscience*, **5**(11), 783-787.
- Riese, M., Ploeger, F., Rap, A., Vogel, B., Konopka, P., Dameris, M. and Forster, P. (2012). Impact of uncertainties in atmospheric mixing on simulated UTLS composition and related radiative effects. *Journal of Geophysical Research: Atmospheres*, **117**(D16).
- Rodwell, M. J. and Jung, T. (2008). Understanding the local and global impacts of model physics changes: An aerosol example. *Quarterly Journal of the Royal Meteorological Society*, **134**(635), 1479-1497.
- Sassi, F., Boville, B. A., Kinnison, D. and Garcia, R. R. (2005). The effects of interactive ozone chemistry on simulations of the middle atmosphere. *Geophysical Research Letters*, **32**(7).
- Semane, N., Peuch, V. H., Pradier, S., Desroziers, G., Amraoui, L. E., Brousseau, P. et al. (2009). On the extraction of wind information from the assimilation of ozone profiles in Météo–France 4-D-Var operational NWP suite. *Atmospheric Chemistry and Physics*, **9**(14), 4855-4867.
- Sigmond, M., Scinocca, J. F., Kharin, V. V. and Shepherd, T. G. (2013). Enhanced seasonal forecast skill following stratospheric sudden warmings. *Nature Geoscience*, **6**, 98-102.

Shepherd, T. G. (2002). Issues in stratosphere-troposphere coupling. *Journal of the Meteorological Society of Japan. Ser. II*, **80**(4B), 769-792.

Tao, M., Konopka, P., Ploeger, F., Grooß, J. U., Müller, R., Volk, C. M. and Riese, M. (2015). Impact of the 2009 major sudden stratospheric warming on the composition of the stratosphere. *Atmospheric Chemistry and Physics*, **15**(15), 8695-8715.

Thompson, D. W. and Solomon, S. (2002). Interpretation of recent Southern Hemisphere climate change. *Science*, **296**(5569), 895-899.

Thompson, A. M., Miller, S. K., Tilmes, S., Kollonige, D.W., Witte, J. C. Oltmans, S. J. et al. (2012). Southern Hemisphere Additional Ozonesondes (SHADOZ) ozone climatology (2005–2009): Tropospheric and tropical tropopause layer (TTL) profiles with comparisons to OMI-based ozone products, *Journal of Geophysical Research: Atmospheres*, **117**, D23301.

Unger, D. A. (1985). A method to estimate the continuous ranked probability score. Preprints. In *Ninth Conf. on Probability and Statistics in Atmospheric Sciences* (pp. 206-213).

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences* (Vol. 100). Academic press.

Williams, R. S., Hegglin, M. I., Jöckel, P., Hella, G. and Shine, K. P. (2021). Composition feedbacks on air quality and weather instigated by sudden stratospheric warming. In Prep.