TECHNICAL MEMORANDUM

704

Model Cycle 38r2: Components and Performance

Peter Bauer, Anton Beljaars, Maike Ahlgrimm, Peter Bechtold, Jean-Raymond Bidlot, Massimo Bonavita, Alessio Bozzo, Richard Forbes, Elias Hólm, Martin Leutbecher, Philippe Lopez, Linus Magnusson, Fernando Prates, Mark Rodwell, Irina Sandu, Agathe Untch, Frédéric Vitart

Research Department

July 2013

This paper has not been published and should be regarded as an Internal Report from ECMWF. Permission to quote from it should be obtained from the ECMWF.

European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen terme

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 2013

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

Literary and scientific copyrights belong to ECMWF and are reserved in all countries. This publication is not to be reprinted or translated in whole or in part without the written permission of the Director-General. Appropriate non-commercial use will normally be granted under the condition that reference is made to ECMWF.

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

On 25 June 2013, a new model cycle (38r2) has been introduced at ECMWF. The main contributions are an increase of the number of vertical levels from 91 to 137 in high-resolution forecast model (HRES), the ensemble of data assimilations (EDA), the main assimilation (4DVAR) and the Boundary-Conditions (BC) optional programme, revised background error variances at 137 levels based on IFS cycle 38r1, revised EDA calibration and filtering for 137 levels, deactivation of the model error cycling in the stratosphere, activation of unbalanced control vector components, a modification of surface drag, test parcel entrainment in boundary layer and shallow convection, autoconversion in convection, an adjustment of non-orographic gravity wave drag to be consistent with System-4, an oxygen absorption correction, and a revision of sea-ice/sea-surface-temperature quality control over the Caspian Sea. The scaling factor for singular vector initial perturbations in the ensemble system (ENS) has been slightly increased.

Tropospheric upper-air scores are overall slightly positive in northern hemisphere and mainly neutral for Europe and southern hemisphere. Performance in the tropics is mixed with some negative results compared to observations but neutral against analyses. In the extra-tropics the main negative impacts are for the upper-tropospheric relative humidity (300 hPa). The main positive impacts are for geopotential in the lower stratosphere, and to a lesser extent also in the troposphere. For precipitation and temperature the overall conclusion is a slight improvement in the extra-tropics and a slight degradation in the tropics.

The scores for 10-metre wind show neutral to slightly positive impact both in extra-tropics and tropics. There is an overall slight reduction of wind speed, most notable in Europe at 12 UTC. No significant differences have been found between the synoptic performance of cycle 38r2 compared to cycle 38r1. Tropical cyclone tracks and intensity have been compared for all Tropical Cyclones available from the 38r2 e-suites. There is a slight improvement for the position errors from day-3 onwards, although this is not statistically significant. The impact is neutral for tropical cyclone intensity.

In general cycle 38r2 is neutral for the ENS in terms of the spread and ensemble-mean skill for the extra-tropics. In the tropics, spread is also changed little. Ensemble mean error is improved for 850 hPa temperature, and slightly worse for 850 hPa wind speed. In terms of probabilistic scores, the CRPS is neutral for 500 hPa height, 850 hPa temperature and wind speed over Europe and the northern extra-tropics. Over the southern extra-tropics e-suite scores are better for 500 hPa height, but worse for days 1-3 for temperature at 850 hPa. In the tropics the e-suite shows a significant improvement throughout the forecast range for 850 hPa temperature, and a small but statistically significant degradation for 850 hPa wind speed, consistent with the changes in the ensemble mean error and HRES.

Contents

1 Introduction

With model cycle 38r2 of the ECMWF Integrated Forecasting System (IFS) a change of the vertical resolution that is employed in the high-resolution model (HRES), the four-dimensional variational assimilation (4DVAR) the ensemble of data assimilations (EDA) and the Boundary Condition (BC) suite has been introduced. The upgrade also affects the ensemble forecasts (ENS) because of its impact on mean model state and variability as well as the non-identical level definition between the initial conditions provided by the EDA and the ENS vertical resolution. The impact on mean model state can also affect the extended, e.g. monthly forecast range.

In January 2010, the horizontal resolution of the high resolution forecasting system had been already increased from T799 (with a 25 km grid) to T1279 (with a 16 km grid). To be able to resolve a similar ratio of vertical and horizontal scales, it was necessary to increase the number of vertical levels by 50%.

Table 1: Past HRES model cycles with vertical resolution upgrades (upgrade was performed with the same cycle).*

Date	Cycle	Model top (hPa)	Number of vertical levels	Horizontal wavenumber truncation	Main other changes*
June 2013	38r2	0.01	137	T1279	Test-parcel in shallow convection, boundary layer, oxygen absorption fix, sub-grid orography, non-orographic gravity wave drag, Jb and EDA recalculation.
February 2006	30r1	0.01*	91	T799*	4DVAR T95/255/799, grid-point humidity and ozone in 4DVAR, revised ozone chemistry, wave model resolution 0.36 deg., 24 directions, 30 wave numbers, Jason altimeter wave heights, ENVISAT ASAR spectra in wave assimilation, ENS resolution T399L62 with model top at 5 hPa.
October 1999	21r4	0.1	60	T319	New orography and climatologies, revised cloud and convection, sub-grid orography and surface albedo treatment in radiation, SSM/I winds active, revised SSM/I and TOVS bias correction, new Jb, ENS L40.
March 1999	19r2	0.1	50	T319	RTTOV-5 radiative transfer model, methane oxidation and photolysis on water vapour in stratosphere, moist adjustment in linear physics, use 10m winds in wave model, introduction of Rayleigh friction in Eulerian advection scheme, modified FULLPOS interpolation of specific humidity in stratosphere.

Figure 1 shows layer depth as a function of height (both in hPa) for L50, L60, L91, L137 versions of the IFS. On the last three occasions, vertical level upgrades have introduced about 50% more levels. Layer depth has been changed throughout the atmosphere but with emphasis on enhancing resolution in the lower troposphere between L50 and L60, upper troposhere and stratosphere between L60 and L91 (also raising the model top) and rather uniformly between L91 and L137. For ENS, the same levels as HRES are maintained in the troposphere and total level number is reduced by lowering the model top. This is expected to change with model cycle 39r1 for which the L91 configuration will be chosen.

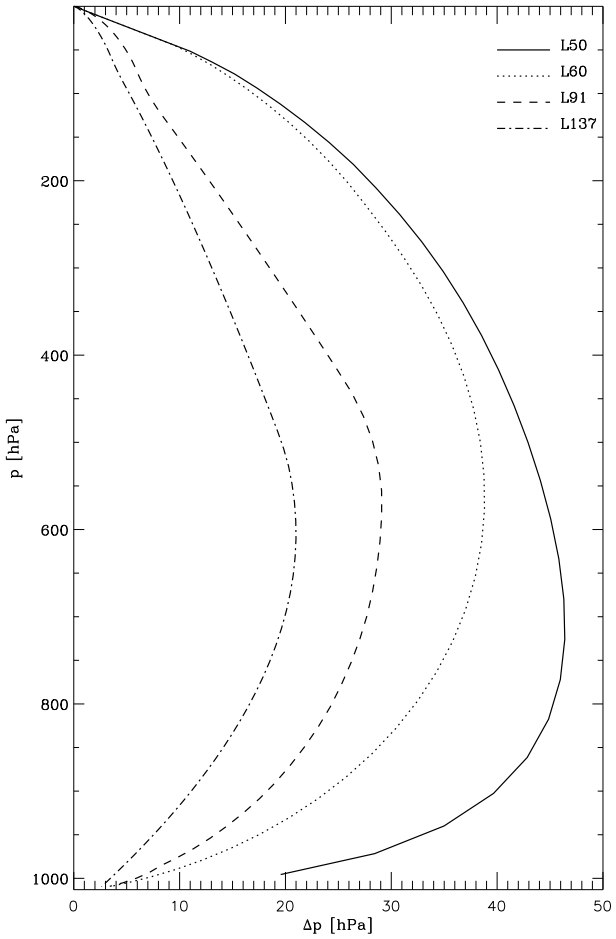


Figure 1: Layer depth as a function of height (both in hPa) for L50 (solid), L60 (dotted), L91 (dashed), L137 (dash-dotted).

Over the past 15 years, several changes of vertical resolution and model top pressure have been introduced in the IFS, as listed together with other changes in Table 1. Note that the IFS was introduced with cycle 11r7 in March 1994. In the short-term, vertical resolution increase tends to be less beneficial in terms of forecast skill than horizontal resolution upgrades. The latter has a straight forward impact on the representation of orography and convection and thus critical near-surface weather parameters as 2-metre temperature, 10-metre wind and precipitation are represented with more detail and extreme weather systems become more realistic in terms of intensity distribution and spatial patterns. Vertical resolution increase is most effective where vertical gradients of temperature, winds (shear) and moisture (clouds) are large, i.e. near the surface across the boundary layer, across the freezing level (melting), in frontal zones, and in the upper troposphere - lower stratosphere (UTLS). Systematic model errors are most pronounced in these areas and change structure with resolution. This affects stability, moist processes through interaction with dynamics and radiation, and vertically propagating waves. Even small changes in mean state and mean model errors as well as enhanced (realistic) variability from increased resolution can produce negative skill compared to smoother structures in a less well resolved model. In addition, hard coded model-level dependent thresholds can produce artefacts once resolution is changed. This is also why horizontal resolution is usually upgraded in 5-year cycles while vertical resolution changes are performed less frequently.

The choice of vertical resolution is by no means obvious. From the purely dynamical point of view, [10] argue that most atmospheric models have far too little vertical resolution. From simple scaling of the quasi-geostrophic equations, they suggest that a typical ratio of vertical to horizontal resolution of $3.5\,10^{-3}$ is needed. They also point out that high vertical resolution is needed for the proper representation of gravity waves in layers where the phase speed of the waves is equal to the wind speed. Earlier vertical resolution changes at ECMWF have always been difficult partially because some of the parametrizations have a rather strong vertical resolution dependence, and secondly because vertical resolution interacts with data assimilation. Many of the resolution dependencies have been deleted or improved over time, particularly in the cloud and convection schemes. Impact in the past could mainly

be seen in the fit to radiosondes because vertical structures could be better captured at high vertical resolution [7]. It is also clear that numerical convergence is very difficult to achieve, not in the least because the real atmosphere has a lot of vertical structure in clouds and inversions which are only controlled by sub-grid parametrizations that tend to be optimized in the context of a particular resolution.

Stable boundary layers are shallow with typical depths of 100m to 200m, but tend to be deeper in models due to rather strong parametrized diffusion which is necessary for large-scale performance reasons [16]. The sensitivity to vertical resolution is surprisingly small mainly because the surface exchange is formulated by prescribing the exact shape of the wind, temperature and moisture profiles between the lowest model level and the surface as part of the numerical formulation [2].

Inversion levels are much more sensitive, because with current resolution of a few hundred metres, a steep inversion of e.g. 10K tends to occur in a single layer. Although these inversions are often very realistic [18], the height of the inversion level can only be discrete, i.e. not more accurate than the model's vertical resolution. Although advanced inversion reconstruction techniques have been developed [11, 5], these techniques are complicated and it is not entirely clear whether they actually manage to improve the numerical accuracy. Previous vertical resolution experiments have always shown that better resolution makes boundary layer growth more smooth by having smaller jumps from one level to the next [2, 19]. Inversions are often cloud boundaries, so having a finer grid to locate an inversion will automatically improve the associated cloud tops. This also applies to cloud base which can only be numerically determined within the resolution of the vertical grid.

It was attempted with cycle 38r2 to reduce the number of modifications to those that address the impact of the vertical resolution upgrade. As can be seen in Tab. 1, this has not always been the case in the past. Particularly with cycle 31r1 a large number of significant changes of vertical and horizontal resolution, position of model top, and data assimilation configuration have been combined so that little information on the individual effect from each contribution is available.

This report describes the initial impact of the resolution upgrade in Section 2, the eventual set of changes applied to the model (Section 3) and data assimilation (Section 4), and the final evaluation of the combined contributions at full model resolution and including both HRES and ENS (Section 5). The latter section also includes examples of areas where immediate benefits from a more detailed description through resolution of physical processes were found, even if these have minor impact on aggregated forecast skill scores. The memorandum is concluded by a brief summary and a list of open issues that need to be addressed in forthcoming model cycles (Section 6) .

2 Initial assessment

2.1 Climate integrations

An ensemble of four 13 month forecasts with cycle 38r1 and 137 levels at T159 resolution were performed to assess the impact of vertical resolution on the 'climate' of the model. The four forecasts are averaged together, seasonal and annual means of different model fields are then calculated and compared to equivalent datasets from satellite observations and the reanalysis (ERA-Interim) for temperature, humidity, winds, clouds, precipitation and radiation. The overall conclusion is that the model 'climate' is very similar for both the L91 and L137 vertical resolution models. However, there are some relatively small systematic effects of vertical resolution that are worth documenting.

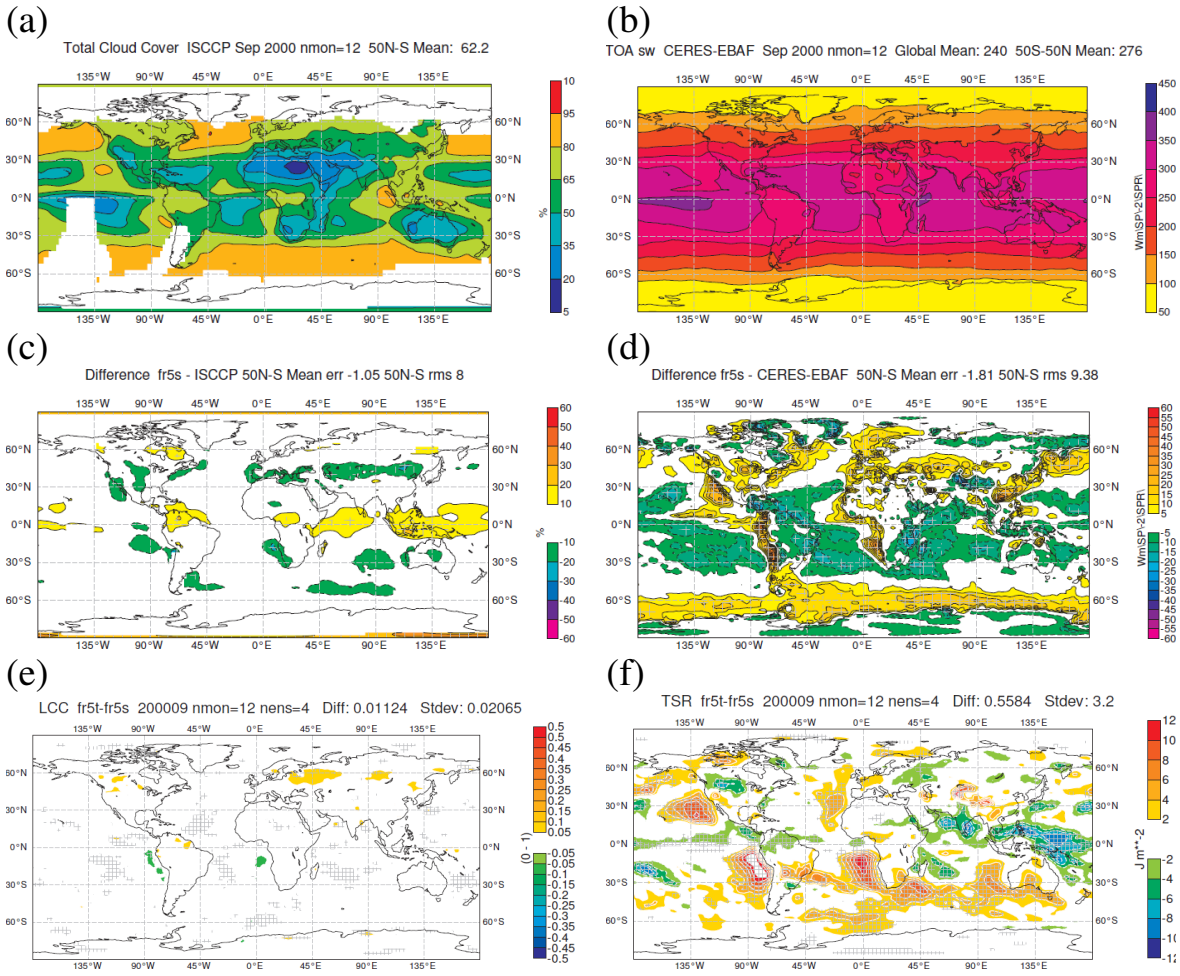


Figure 2: Annual mean differences for one year T159 model integrations versus observational datasets for cloud cover and top-of-atmosphere net solar radiation. (a) ISCCP total cloud cover (%), (c) L91 model minus ISCCP total cloud cover(%), (e) L137 minus L91 low cloud cover (0-1), (b) CERES-EBAF TOA net shortwave radiation (W m^{-2}), (d) L91 model minus CERES-EBAF (W m^{-2}), (f) L137 minus L91 (W m^{-2}). Note the colour scales are different for each plot.

Figure 2(a) shows the annual mean total cloud cover from the ISCCP dataset and Fig. 2(c) shows the difference between the L91 model and ISCCP. There is a general overestimate of cloud in the reference model associated with convection in the Tropics, an underestimate in mid-latitudes and an underestimate

in the marine stratocumulus regions in the south Atlantic and Pacific. The pattern of the differences is similar versus other satellite cloud datasets so there is some confidence that these are systematic errors in the model. The impact of the vertical resolution increase is shown in Fig. 2(e) highlighting a small (5%) increase in cloud cover over northern hemisphere mid-latitude land, but a 5% decrease in cloud cover in the marine stratocumulus regions. The latter leads to a slightly worse agreement with the satellite datasets. The cloud liquid water associated with the marine stratocumulus is also reduced (not shown) and the two effects combined have an impact on the reflection of solar radiation. The equivalent plots for the top-of-atmosphere net shortwave radiation are in the right-hand column of Fig. 2 for (b) the annual mean from the CERES-EBAF satellite dataset, (d) the difference of the reference L91 model from the CERES-EBAF observations and (f) the difference of L137 minus L91. The impact of increasing vertical resolution on the top-of-atmosphere net shortwave radiation particularly highlights the higher values (less outgoing radiation) due to the decreased low cloud cover and cloud water content in the marine stratocumulus regions. There is also a smaller but opposite bias in the regions of deep convection in the tropics due to an increase in liquid water content.

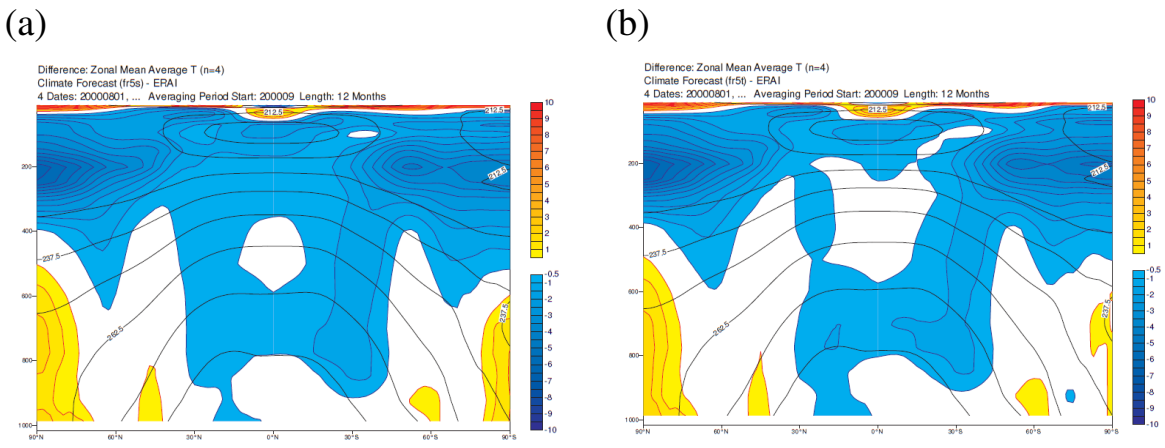


Figure 3: Annual mean zonally averaged temperature difference (K) from ERA-Interim for one year T159 model integrations for (a) L91 and (b) L137 vertical resolutions

The higher vertical resolution also has a small impact on the tropospheric temperature and humidity in the tropics. Figure 3 shows the annual mean zonally averaged temperature difference of the L91 and L137 models from ERA-Interim. The L137 model is slightly colder (0.3 K) in the tropics between 800 hPa and 600 hPa (worse compared to ERAI) and slightly warmer (0.3K) between 600hPa and 200hPa (in better agreement with ERAI). The humidity in the lower troposphere is also slightly higher (not shown). In addition, there is a 5 % decrease in convective precipitation and a small increase in stratiform precipitation in the L137 model. This suggests the formation rate of convective precipitation is reduced for higher vertical resolution. The changes in temperature and humidity in the lower troposphere and the convective precipitation may be related and this is followed up in the next section on medium-range forecasts and process tendencies.

2.2 Medium-range forecasts and process tendencies

The impact of the vertical resolution change is investigated for a series of ten day forecasts to determine how the differences evolve in the first few days and to help identify the processes involved. Averages are produced over a number of forecasts for both the L91 and L137 configurations of cycle 38r1 starting from identical initial conditions. A number of different configurations were tried, including different

resolutions (T159, T511, T1279), different initial analyses (from L91 and L137 analysis experiments), different seasons and averages over different numbers of forecasts. The differences between the L137 and L91 models were found to be robust to different vertical resolution analyses and to different horizontal resolutions of the forecast model so most experimentation could be performed at lower resolution and the T159 results are shown here. An averaging period of 15 days was sufficient to reduce noise and highlight the main signals (i.e. averages are over 15 forecasts initialised each day at 00Z for 01-15 January 2012).

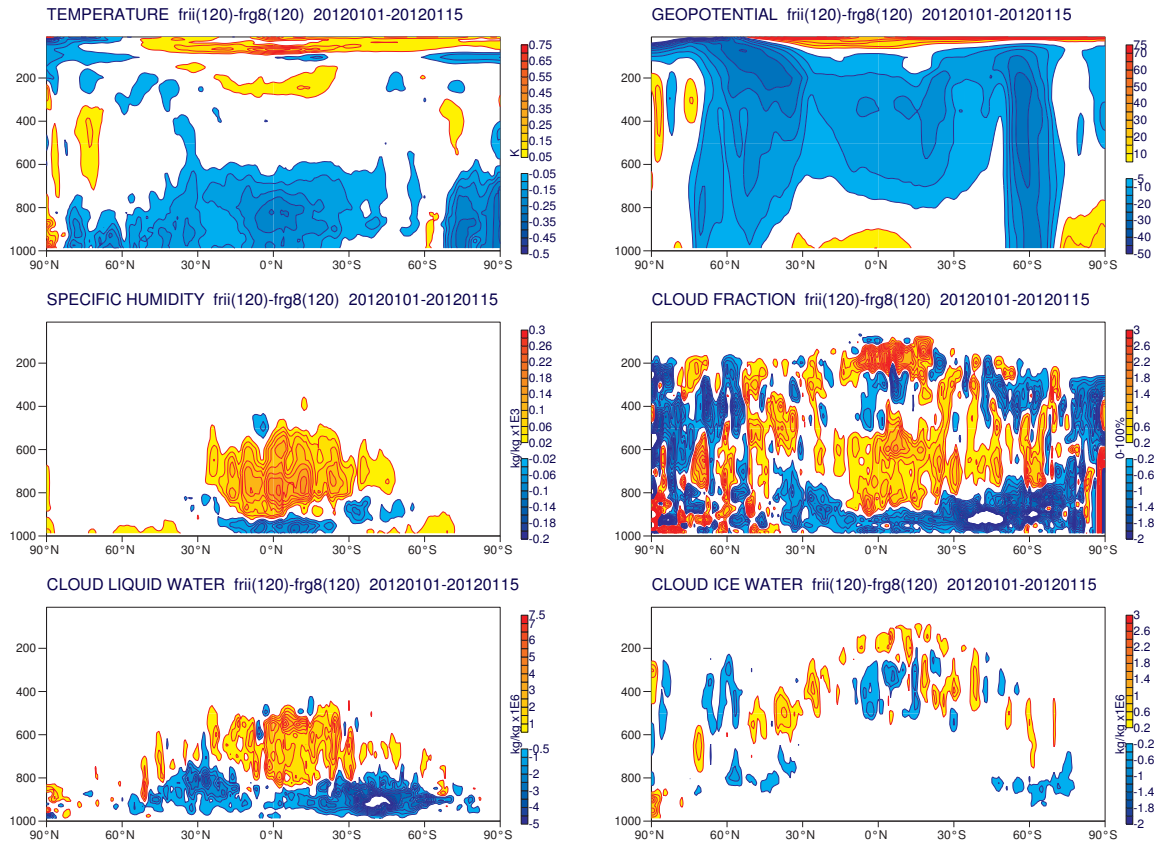


Figure 4: Differences of L137 minus L91 day-5 forecasts averaged over 15 forecasts from 1-15 January 2012 for temperature (K), geopotential height (m), specific humidity (10^3 kg kg^{-1}), cloud fraction (%), cloud liquid water and cloud ice water (both 10^6 kg kg^{-1}). Note, the cross-section is not masked for orography.

Figure 4 shows cross-sections of the zonally averaged differences in temperature, geopotential, humidity and cloud fields on pressure levels 5 days into the forecast. The L137 model first becomes colder below 600 hPa in the tropics by about 0.25 K (as well as some parts of the extra-tropics) and gradually warms in the upper half of the tropical troposphere (and lower stratosphere) increasing with lead time. This affects the geopotential height with lower values higher in the troposphere as a result of the integrated cold temperature anomaly. The humidity is increased in the tropics below about 500 hPa, and decreased in a thin layer near the surface. This corresponds with an increase in tropical cloud fraction and cloud liquid water between 900 and 500 hPa and a reduction in both below. At high altitudes in the tropics there is also increased cloud fraction and some small changes to the vertical distribution of cloud ice water. The reduced cloud cover and cloud liquid water from boundary layer cloud coming from the marine stratocumulus regions and southern hemisphere are also evident.

To understand which processes might be the source of the differences, the tendencies from the dynamics and each of the parametrizations are extracted and averaged as described above. Only the temperature

and humidity tendencies are discussed here. Figure 5 shows the temperature difference 5 days into the forecast and the contributions from the different temperature tendencies from the dynamics, radiation, vertical diffusion, convection and stratiform cloud scheme averaged over the 5 day period. Firstly, there is a significant balance of tendencies with different processes dominating in different regions. The main lower tropospheric cooling signal is from the convection and cloud scheme, whereas the upper tropospheric warming in the tropics is primarily from the convection. There is a response at the melting layer in the tropics in the convection, cloud scheme, dynamics and radiation. This is very sensitive to small changes in the amount of convective and stratiform snowfall reaching the zero degree isotherm and is also a result of the higher resolution in the L137 model slightly changing the height of the melting level.

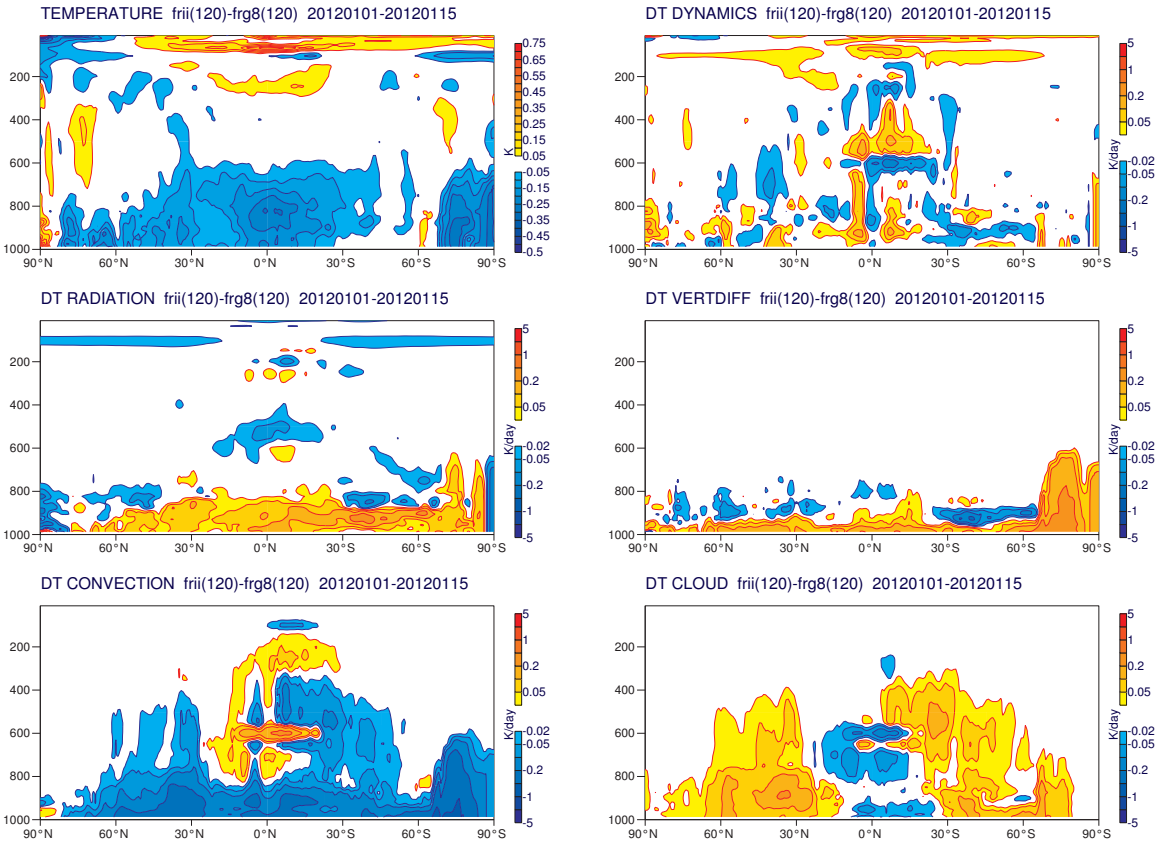


Figure 5: As for Fig. 4 for temperature difference (L137 minus L91, top left) after 5 days and temperature tendencies (DT; K/day^{-1}) averaged over the 5 days from dynamics, radiation, vertical diffusion, convection and the stratiform cloud scheme. Note the non-linear scale for the temperature tendencies to highlight both small and larger changes.

The vertical diffusion is predominantly warming (and drying) in the boundary layer. The radiative heating is responding to the changes in humidity and cloud in the lower troposphere. The narrow band of cooling at 100 hPa in the radiation tendencies is due to a sharp transition of gaseous assumptions in the computation of the transmission function between the troposphere and stratosphere defined at this level and is discussed in more detail later in this report. The increased humidity in the tropical lower troposphere seen in Fig. 4 is also mainly from the convection and cloud schemes, with the drying in the boundary layer originating from the vertical diffusion (not shown).

Figure 6 shows the impact of the vertical resolution change on top-of-atmosphere net solar radiation and convective precipitation averaged over the 5-day forecasts. The differences correspond well with those

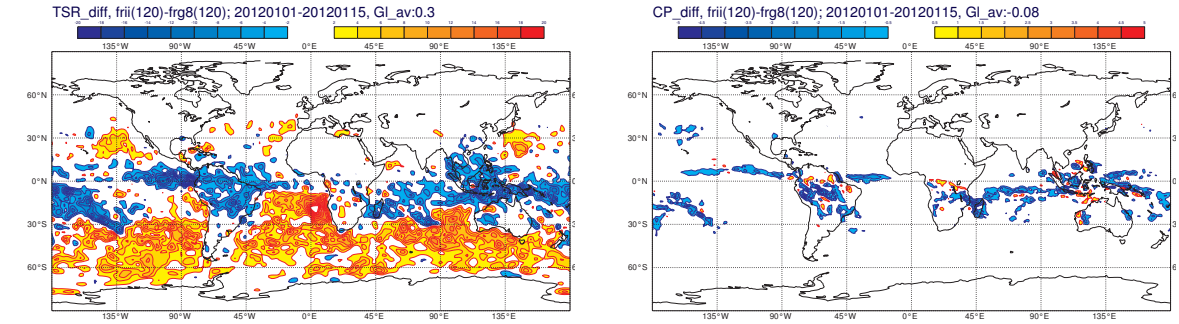


Figure 6: Differences of L137 minus L91 day-5 forecasts averaged over 15 forecasts from 01-15 January 2012, (a) top-of-atmosphere net solar radiation (W m^{-2}), (b) convective precipitation (mm day^{-1}).

seen in the climate simulations (Fig. 2) with decreased reflection in the marine stratocumulus regions, sub-tropics and mid-latitudes, and a decrease in net solar radiation in regions of deeper convection in the tropics. There is also a reduction in the convective precipitation associated with deep convection of about 5%.

3 Model changes

Although the overall differences between the two vertical resolution models are small, there are a number of systematic changes that are consistent both in the short-range, medium-range and climate of the model. These changes are likely to be a result of many different sensitivities to vertical resolution in the dynamics and in the parametrization schemes which are not straight forward to identify or solve. However, some of the systematic differences can be reduced by targeted changes to the parametrization schemes. In particular, the decrease in marine stratocumulus is addressed with a change to the test parcel entrainment in the convection and cloudy boundary layer schemes which improves the consistency of the two schemes and rebalances the occurrence of overcast stratocumulus and partially cloudy shallow convection. The source of the radiation difference at 100 hPa is identified and will be addressed in the future with smoother transmission functions between troposphere and stratosphere. The lower tropospheric cooling and moistening in the tropics and decreased convective precipitation can be addressed by modifying the convective rain formation. These changes are all described in more detail in the following sections.

3.1 Sub-grid orography scheme

Initial comparisons of the surface fields of forecasts running cycle 38r1 with the new L137 vertical resolution to those with the L91 resolution revealed a small systematic difference in the surface stress produced by the sub-grid orography (SO) scheme. The SO stress with the L137 model is always lower than with the L91 model. The magnitude of the difference is between 10% and 15% which is clearly seen from the zonal mean plots of Figure 7a.

The main underlying cause of the resolution dependence is in a level searching algorithm. The first routine that is called in the subgrid orography scheme defines a number of levels by indices that are later used to control the depth over which blocking and gravity wave generation occurs. Three of such indices are considered, which are defined by the model level index where the height above the surface reaches: 3 times the standard deviation of the subgrid orography ($3\mu_{so}$), μ_{so}, and the level where the vertical integral starting from $3\mu_{so}$ of N/U (Brunt-Vaisala frequency divided by wind) reaches a critical Froude number of 0.5. The latter is called the blocking height. Below the blocking height, the flow is assumed to go around the orography and is exposed to a strong drag, whereas the flow above the blocking height is assumed to go over the orography generating gravity waves [12].

By construction these heights are discrete and therefore the result depends on vertical resolution. The level searching algorithm scans from the surface upwards and defines the level index as soon as the requested height is reached at a full level. This results in systematic differences when resolution is increased. Searching for the nearest level (which implies that the height criterion is applied to half levels), leads to random differences when resolution is changed, but should not show systematic differences. We will call the latter algorithm half level searching (HLS).

The effect of vertical resolution and HLS on the mean of $3\mu_{so}$, μ_{so}, and the blocking height, as projected on the vertical grid, has been checked and it is clear that the control model has a systematic sensitivity to vertical resolution which is eliminated with HLS. The zonal means of the discrete representation of these heights are in fact identical with HSL (not shown). The surface stresses with HSL for the two resolutions is shown in Fig. 7a. HSL does indeed reduce the sensitivity of surface stress to vertical resolution but does not eliminate it completely. The reason for the remaining resolution dependence is not clear. The change to HLS also changes the stress systematically, because it lowers the heights on the average by

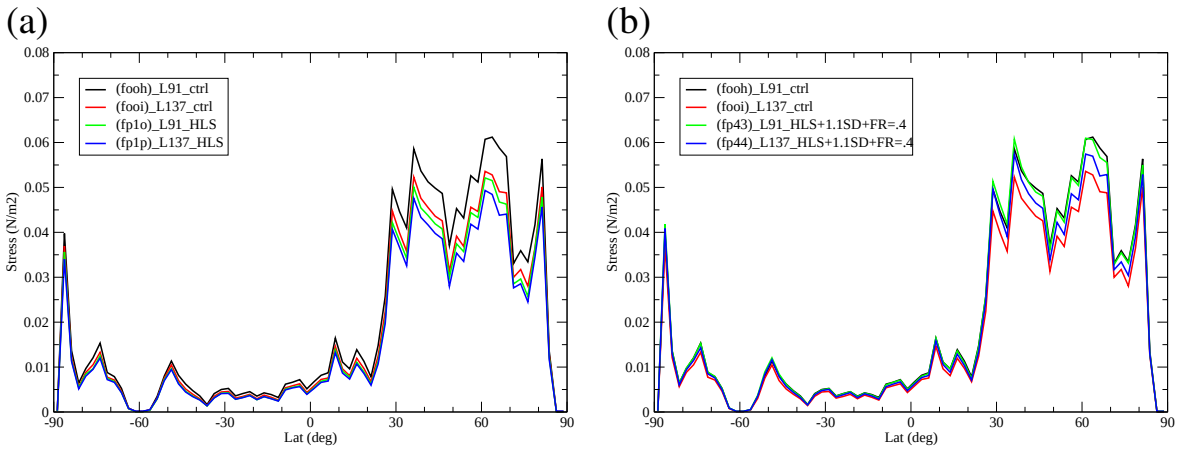


Figure 7: Zonal mean average of the subgrid orography stress over the first 24 hours of a T511 forecast from 20120101 with the L91 (black) and the L137 (red) models. The green and the blue curves are with half level searching (HLS) at the two resolutions (a) and with HLS and parameter adjustments (HLS-ADJ) (b).

half a model layer thickness.

Although the modified algorithm alleviates the resolution dependence, the surface stress is lower than in the control model. In order to bring the SO-stress back to the level of the control model (cycle 38r1), adjustments are necessary. There is a range of candidate parameters. In this case, parameters are selected that are related to the height scale definition. Therefore the levels that control the height range over which blocking and gravity wave generation are active are increased by a factor 1.1, i.e. the levels μ_{so}, $2\mu_{so}$, $3\mu_{so}$, and $4\mu_{so}$ are replaced by $1.1\mu_{so}$, $2.2\mu_{so}$, $3.3\mu_{so}$, and $4.4\mu_{so}$. Furthermore, the critical Froude number that is used for the blocking height is reduced from 0.5 to 0.4.

The impact on the surface stress is shown in Fig. 7b. The surface stress is close to the control model but not all areas behave in the same way. The stress is still a little below the level of model cycle 38r1. To have a preliminary evaluation of the modified subgrid orography scheme, forecast only experiments were run for January 2012 and July 2011. The impact of HLS-ADJ on the Northern Hemisphere 500 hPa forecast scores is very small at both resolutions (not shown) although a small sensitivity can be seen for 100 hPa winds (not shown). On the basis of the results above it was decided to include HLS-ADJ in cycle 38r2. The changes reduce the dependency of the subgrid orography scheme on vertical resolution and have very little impact on the large-scale model performance.

3.2 Boundary layer and shallow cumulus

In boundary layer as well as shallow convection schemes, a so-called ”test parcel” is used to determine the level at which a parcel of air rising from the surface would stop (i.e. the level of zero buoyancy), which is further needed in the two schemes. One of the most important factors controlling the zero-buoyancy level is the mixing of environmental air into the parcel which is controlled by an entrainment rate.

Experiments with L137 revealed a detrimental reduction of stratocumulus cloud occurrence because the formulations of the entrainment rates were not consistent between both schemes, which was causing problems for their interaction. With the new model cycle 38r2, the same formulation for the entrainment is used for the test parcel in both schemes. The coefficients used in this formulation also became identical. They have been adjusted based on the performance of the model in both climate mode runs and forecast

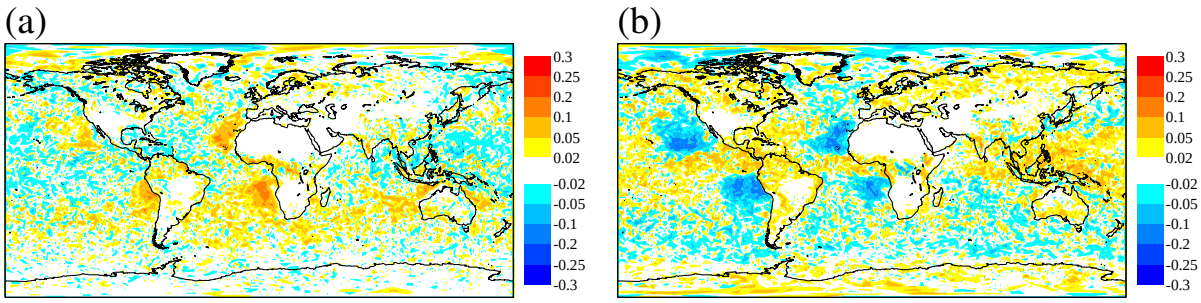


Figure 8: Difference in low cloud cover between cycle 38r2 and 38r1 entrainment formulations run at 137 levels (a), and between 137 and 91 level runs with the cycle 38r1 entrainment formulation (b). Results from 24-hour forecasts averaged over June 2011.

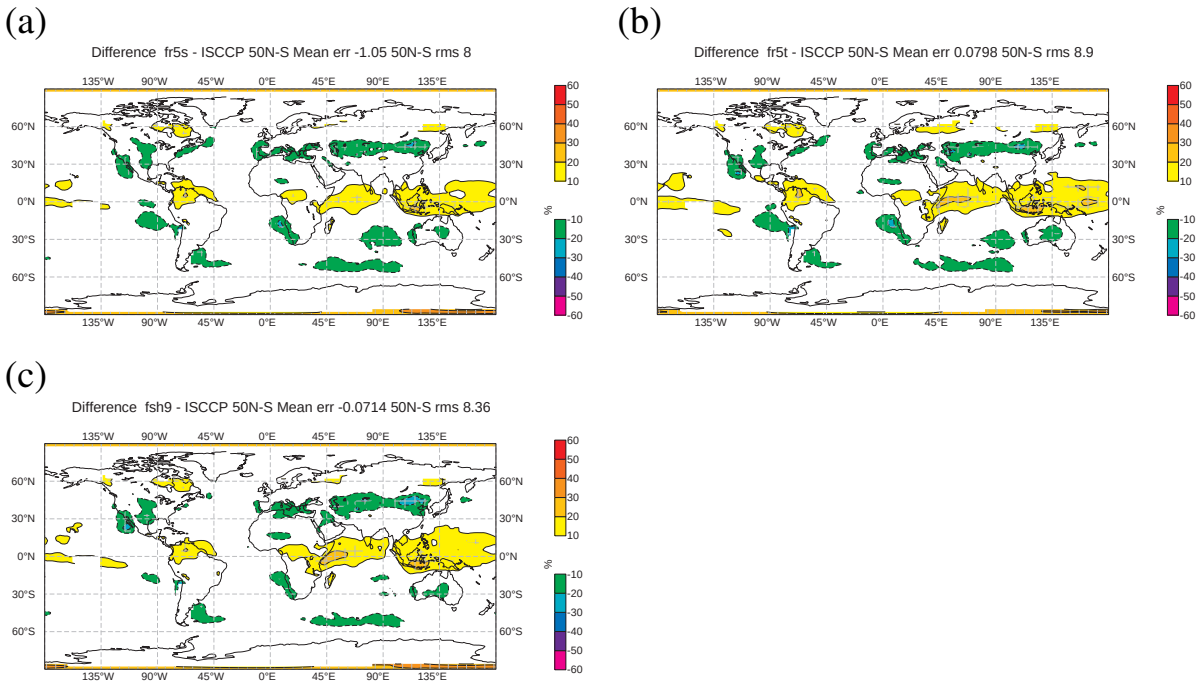


Figure 9: Difference in low cloud cover (%) with respect to ISCCP cloud product from climate runs performed with T511L91 control experiment (a), T511L137 experiment with cycle 38r1 (b) and 38r2 entrainment formulation (c).

runs for both L91 and L137 configurations.

Figure 8a shows the mean June 2011 difference of low cloud cover (LCC) of 24-hour forecasts run with T511L137 experiments using the new and the old entrainment formulations. Both experiments also produced their own analyses which have initialized the forecasts, respectively. The change in LCC is mainly attributed to the change of the entrainment formulations in the two test parcels. The effect is beneficial, in that it partially compensates for the destruction of stratocumulus clouds and the increase in shallow clouds which was introduced by the vertical resolution change. This is shown in Fig. 8b as the difference in LCC between the 137 and 91 runs, both using the old cycle 38r1 formulation. A similar signal has been found for winter (not shown), as well as in climate runs (Figure 9).

3.3 Convection autoconversion

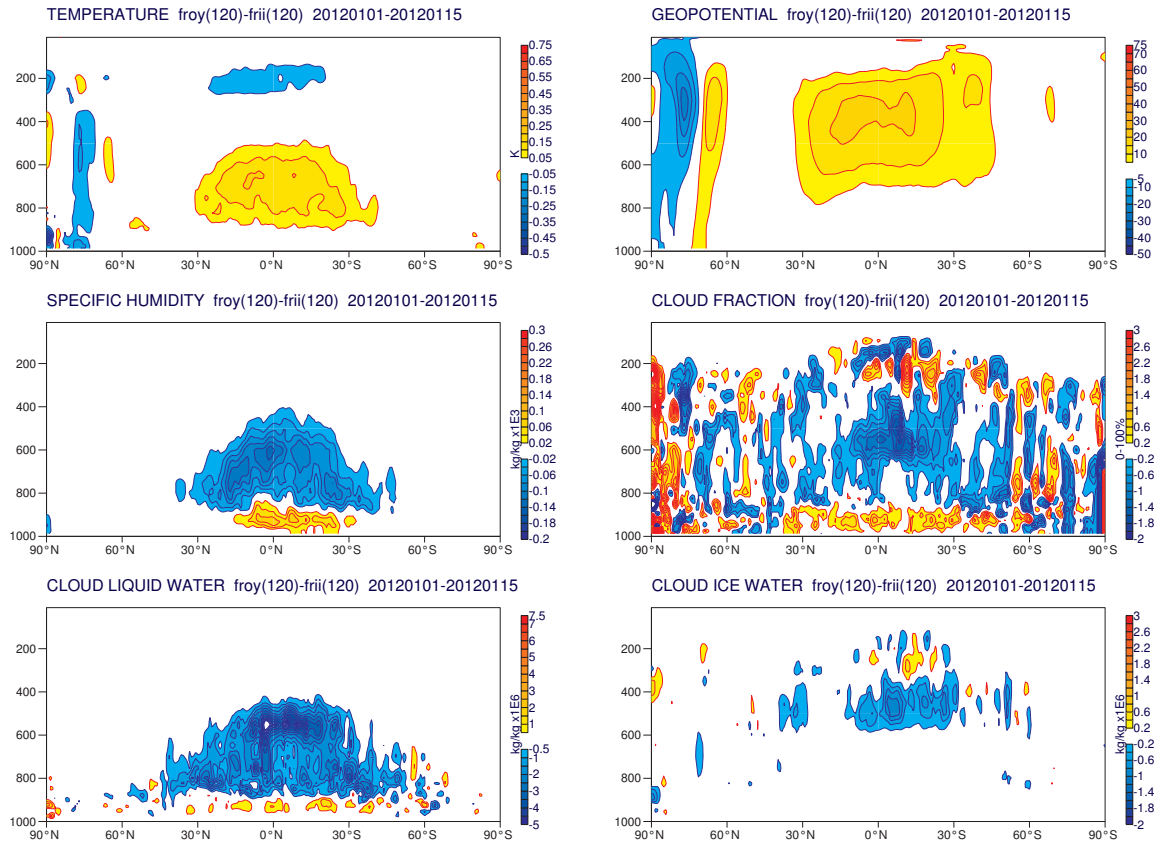


Figure 10: Impact of the increase in warm-phase convective autoconversion rate (L137+autoconvesion minus L137) averaged over 15 forecasts from 01-15 January 2012 for a forecast lead time of five days, for temperature (K), geopotential height (m), specific humidity (10^3 kg kg^{-1}), cloud fraction (%), cloud liquid water and cloud ice water (both 10^6 kg kg^{-1}). To be compared with Figure 4.

The comparison of the L137 and L91 model versions in the one year 'climate' integrations and medium-range forecasts in Section 2.1 show the lower troposphere in the tropics for higher vertical resolution is colder and moister, with increased mid-level cloud fraction and cloud liquid water and reduced convective precipitation by about 5 %. Although the reasons for this sensitivity may be complex dependencies and interaction between schemes, one way to reduce the difference is to increase the rate of formation of convective precipitation. In the IFS, the convection and cloud schemes are two parts of the convection representation. The convection scheme determines the amount of condensation (heating) in convective updraughts above cloud base and detrains humidity, condensate and cloud fraction to the 'stratiform' cloud parametrization scheme. The cloud scheme can then evaporate this detrained cloud (cooling) in the environmental air due to convectively induced subsidence or turbulent erosion. So, an increase in the convective autoconversion rate would increase convective precipitation, taking moisture out of the system and reducing the condensate and cloud fraction passed to the cloud scheme. This would result in less evaporation and therefore less cooling and moistening. If applied to the warm-phase cloud only, the result is a warmer, drier lower troposphere in the tropics with less cloud liquid water and less cloud cover.

Figure 10 shows the impact of a 30 % increase in the autoconversion rate of warm-phase convective

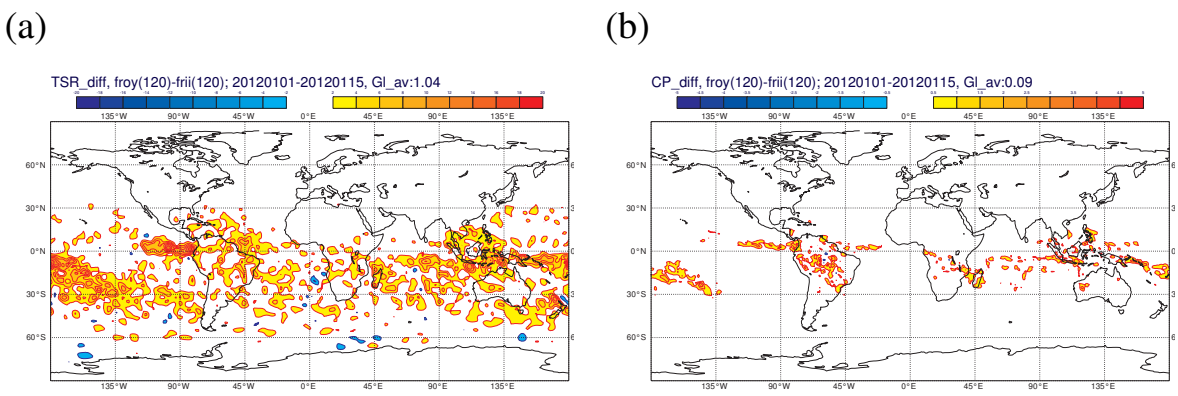


Figure 11: Impact of the increase in warm-phase convective autoconversion rate (L137+autoconv – L137) averaged over 15 forecasts from 1-15 January 2012 for a forecast lead time of five days, (a) top-of-atmosphere net solar radiation (W m^{-2}), (b) convective precipitation (mm day^{-1}). To compare with Fig. 6.

precipitation in the L137 model for a series of 5 day forecasts (equivalent to the first experimentation integrations described in Section 2.1 and to compare with the vertical resolution forecast differences shown in Fig. 4). The change has the desired impact and counteracts some of the main differences seen between the different vertical resolution integrations. In addition, Fig. 11 shows the top-of-atmosphere net solar radiation and convective precipitation (to compare with the vertical resolution forecast differences shown in Fig. 6). The decrease in mid-level cloud and convective precipitation in deep convective regions seen with the increase to 137 levels is also counteracted to some extent by this change. The lower tropospheric warming, reduced cloud cover and increase in top-of-atmosphere solar radiation also improve the model systematic errors compared to observations (climate integrations and medium-range forecasts) at both L137 and the lower L91 vertical resolution.

In summary, although essentially a tuning to the very uncertain process of convective rain formation, an increase of 30 % to the warm-phase convective autoconversion rate mitigates the lower tropospheric temperature and humidity changes seen with increasing vertical resolution and leads to improved temperature, humidity and geopotential height rms-errors in the tropics.

3.4 Non-orographic gravity wave drag and stratospheric winds

The stratospheric circulation is determined by radiative cooling (heating) in the winter (summer) hemisphere and the drag exerted by vertically propagating synoptic and gravity waves. The resolved wave drag largely depends on the horizontal and vertical resolution of the model. It can be shown that the resolved wave drag is larger and more realistic with 137 levels compared to 91 levels, partly because the vertical phase speed of the waves is more realistic. This enabled us to make two important changes to the parametrized subgrid wave drag in this model cycle: 1) Rayleigh friction on momentum is no longer applied (in the L91 version it was applied to wavenumber zero above 5 hPa) and 2) the parametrized non-orographic wave momentum flux at the launch level is no longer globally uniform, but is made latitude dependent, decreasing the flux into the tropics by apllying a function with Gaussian weights.

These modifications have a strong impact on the representation of the quasi-biennal oscillation (QBO) which is characterized by a wind reversal in the tropical stratosphere with a period of about 2 years. Figure 12 displays the 6-year evolution of the tropical winds between 1993 and 1999 in the ERA-Interim reanalysis, and that produced by 6-year integrations at resolution T159 with cycles 38r1 L91, 38r2 L137 without and with the changes in the parametrized wave drag. The wind evolution in the analysis (Fig-

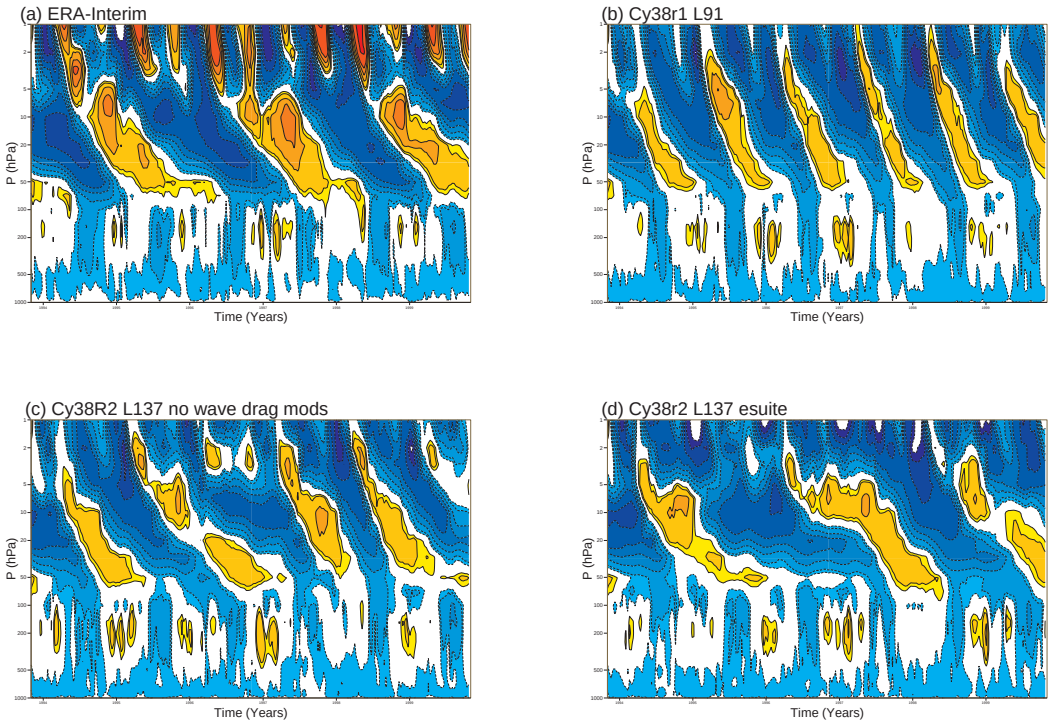


Figure 12: Evolution of the u-wind component averaged over the tropical band during 1993 to 1999 from (a) ERA-Interim, and from 6-year integrations at T159: (b) cycle 38r1 with 91 levels, (c) cycle 38r2 with 137 levels but without gravity wave drag modification, and (d) cycle 38r2 with 137 levels including the gravity wave drag modification.

ure 12a) reveals indeed a downward propagating signal in the stratosphere with a period of about 2 years. Also, near the stratopause at 1 hPa the semi-annual oscillation becomes apparent. Cycle 38r1 with L91 produces a stratospheric wind reversal, but within a much shorter period of roughly one year (Figure 12b). This is clearly improved with the introduction of 137 levels (Figure 12c) which increase the wind reversal period to roughly 1.5 years. However, a too short period is indicative of too large wave drag, and only the introduction of the 137 levels together with the reduction in the parametrized wave drag discussed above, allowed for a realistic representation of the QBO (Figure 12d) in the Cycle 38r2. But the amplitude of the semi-annual oscillation is still too weak and its representation remains challenging.

An evalution of the stratospheric winds in the high-resolution system during January-March 2012 is shown in Figure 13. During this season the mean winds at 30 hPa (Figure 13a) are characterised by mid-latitude westerlies, whereas the tropical winds are in the easterly phase of the QBO. At 50 hPa, however, (Figure 13b) westerly tropical winds prevail along and north of the equator. Differences in rms errors of the day-5 forecasts between cycle 38r2 and 38r1 are shown in Figures 13c-d. There is a notable improvement in 38r2 in tropical regions with westerly winds and along the shear zones of the subtropical westerlies over Central Asia and south of 30° N. Figures 13e-f show the covariance between 30 and 50 hPa winds as computed from the analyses, and the difference in the covariance error between cycle 38r2 and 38r1. The correlation between 30 and 50 hPa winds is close to one beyond 30° latitude but it is either weak or negative in the tropical band. In the forecasts, these correlations increase in the tropical band with forecast time, possibly due to a reduction in vertical shear. Cycle 38r2 maintains a more realistic

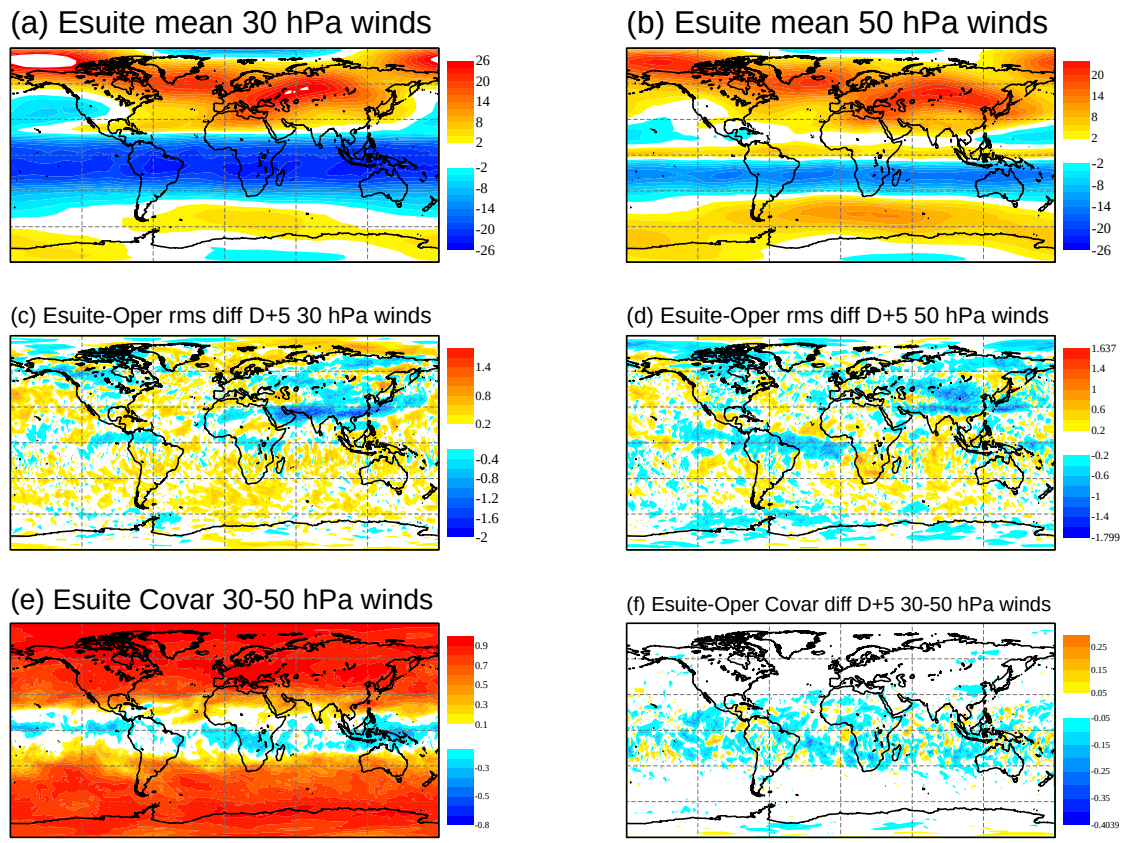


Figure 13: Evaluation of the u-wind component in the stratosphere between January and March 2012: (a)-(b) mean wind at 30 and 50 hPa, (c)-(d) difference in rms error of the day-5 forecasts between cycle 38r2 and 38r1 evaluated, verified with own analyses, (e) covariance between the analysed 30 and 50 hPa winds, and (f) difference in covariance errors between cycle 38r2 and 38r1. Improvements correspond to negative values (blue colours).

wind profile and reduces the error in the (de)coupling between the 30 and 50 hPa winds by roughly 50%. These results seem to be consistent with the more realistic wind evolution seen in Figure 12.

3.5 Radiation

Plots of the zonal mean difference in radiative heating rates between the 91 and 137 level models revealed a band of cooling of the order of 0.02 K/day at 95.583 hPa (Figure 14a). The discontinuity only depends on the long-wave (LW) heating rate. This is not a bug but a feature of the longwave Rapid Radiative Transfer Model (RRTM). As presented in the documentation (Table 2.3, p. 21 of cycle 38r1 documentation), the computation of the transmission function in the different spectral intervals does not use the same gases in the troposphere and stratosphere, with the transition between the two being set at log(pressure) = 4.56. This transition only appears as a minor change in the slope of the LW heating rate between 91 and 137 levels. When plotting the difference between the two vertical resolutions based on pressure levels, the interpolation from model to pressure levels makes this feature visible. As this 'boundary' does not seem to have any impact on the temperature field and convection usually does not cross the 95.583hPa level, this is not considered a problem.

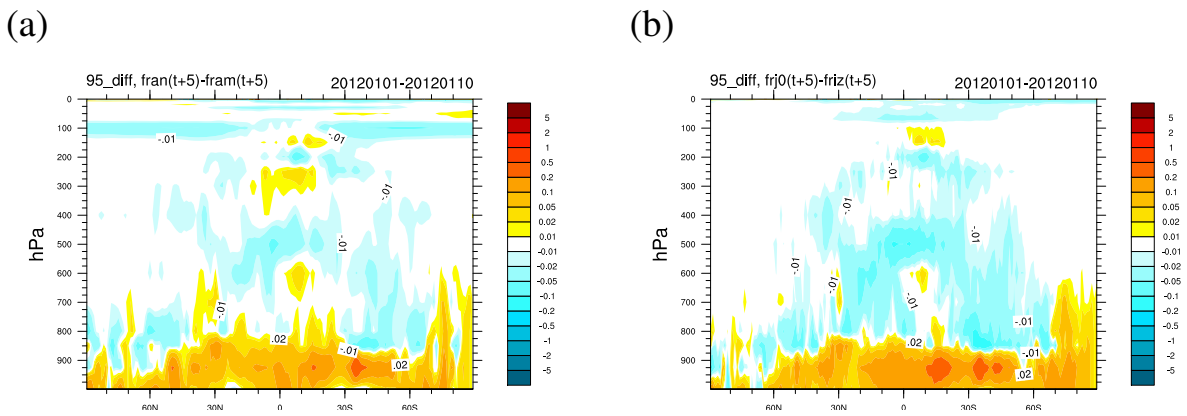


Figure 14: Day-5 net radiative heating rate (K/day) averaged over 10 forecasts. L137-L91 for operational radiation code (a) and with updated spectroscopic database for radiation (b).

The most recent version of longwave RRTM treats the gaseous absorption across the 95.583 hPa level in the same way as the older version that is operational since cycle 32r2 (June 2007), but the updated absorption coefficients produce a smoother profile in the LW fluxes. Preliminary tests show (Fig. 14b) that the updated spectroscopic database eliminates the discontinuity in the longwave heating rate and hence its dependence on vertical resolution.

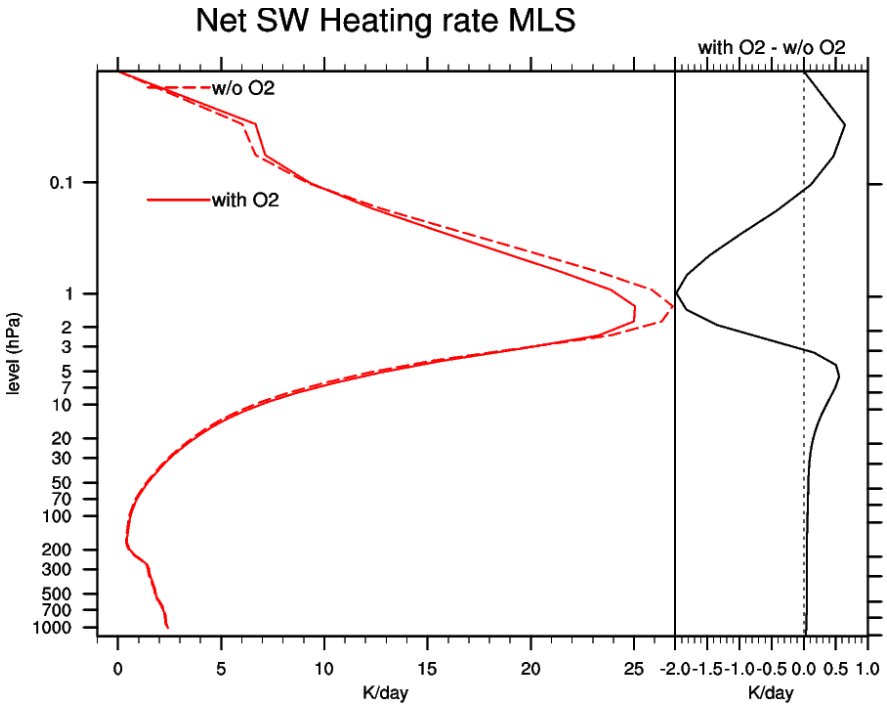


Figure 15: Mid latitude Summer, short-wave heating rate with and without the contribution of O_2 absorption, averaged over a 12-month period September 2000-August 2001.

In addition, a bug was discovered in shortwave RRTM affecting the computation of the clear-sky extinction optical depth in the shortwave spectrum. The column amount of molecular Oxygen (O_2) was set to zero thus neglecting the radiative effect of O_2. The bug affects the short-wave scheme since cycle 32r2 when the shortwave RRTM scheme was introduced. SW O_2 absorption is active in the UV spectrum

(0.15-0.25 μm), in the visible (0.63-0.76 μm) and in the near-IR (1.06-1.58 μm). These absorption lines are included in bands 22, 24, 28 of the shortwave RRTM both in the troposphere and in the stratosphere. The bug has been fixed by setting the volume mixing ratio for O_2 to 0.209488 ppmv (as in US Standard Atmosphere, 1976). Although O_2 plays a minor role in the absorption of shortwave radiation, it is a well mixed gas with homogeneous distribution and the omission of its radiative contribution has a non-negligible impact on SW fluxes and heating rates. The enhanced absorption of solar radiation increases the clear-sky shortwave heating rate mostly in the stratosphere by up to 0.6 K/day around 5hPa but also in the lower-stratosphere and troposphere by about 0.07-0.04 K/day between 50 hPa and 500 hPa (Figure 15).This leads to a temperature increase of 0.5-1.5 K between 2 hPa and 200 hPa, as can be seen from a 12-month model integration (not shown).

Above 2 hPa the solar heating increases towards its maximum due to the combined absorption by O_3 and O_2, mostly in the UV bands between 200 nm and 350 nm. The inclusion of the contribution of O_2 to the total absorption translates in RRTM to a vertical redistribution of the solar heating rate profile with a reduction of the heating around 1hPa and an increase above 0.5hPa and below 3hPa. The net effect is a relative cooling of the upper stratosphere above 2 hPa. A reduction in total downwelling SW flux at the surface is also observed, and this can reach 5 W/m^2 in clear-sky conditions in comparison to an atmosphere without O_2 absorption.

3.6 TL/AD aspects

The experimentation for the initial cycle 38r2, including the changes applied to physical parameterizations asd described above, exhibited the frequent occurrence of very large condition numbers (CN) in all 4DVAR minimizations, which could occasionally lead to a failure of the minimization. A quick work-around was to set the perturbations of all exchange coefficients (both upper-air and surface) to zero in the linearized vertical diffusion parameterizations. This temporary solution was adopted for the final version of cycle 38r2.

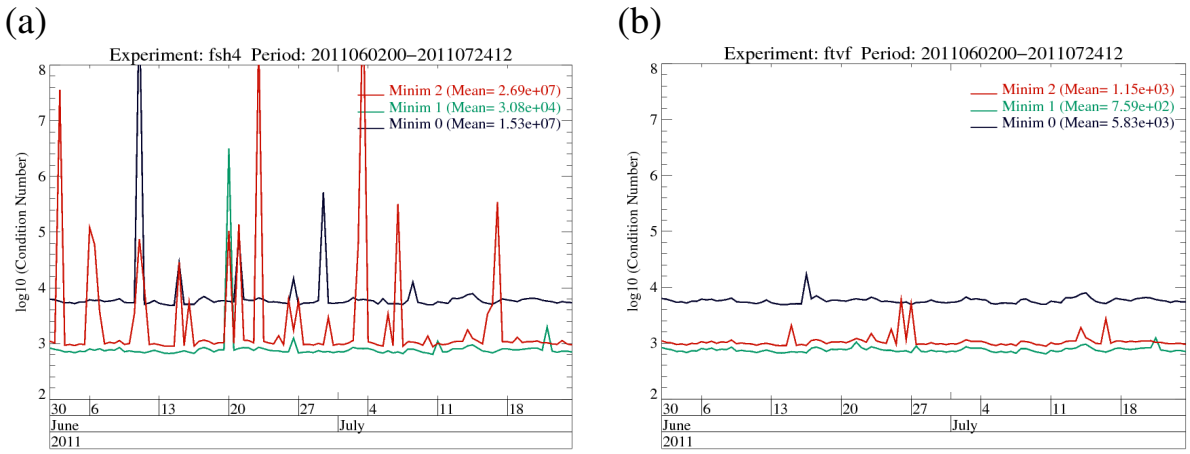


Figure 16: Time series of condition numbers from initial T511 L137 38r2 analysis experiments before (a) and after (b) revision of the regularization of exchange coefficients for momentum in the linearized vertical diffusion.

However, removing all perturbations of the exchange coefficients is not desirable because this results in a degradation of the tangent-linear assumption, i.e. how well the tangent-linear model matches its full non-linear counterpart. It also has the disadvantage of preventing surface perturbations from influencing atmospheric fields during the minimization. Some work was therefore needed to try to identify

more precisely the source of the instability in the linearized physics and to eliminate it, while restoring the inclusion of non-problematic perturbations. It was found that the instability could only arise from the perturbations of the exchange coefficients for momentum (both upper-air and surface). A stronger reduction of momentum exchange coefficient perturbations was therefore coded and tested.

Figure 16a shows an example of the large condition numbers that were found in the initial T511 L137 38r2 experiment in summer 2011. Note that the scale along the y-axis is logarithmic. Normal values for the decimal logarithm of the condition number should fluctuate around 3.8 (i.e. CN≈6000) in the first minimization and below 3 (i.e. CN<1000) in the second and third minimizations.

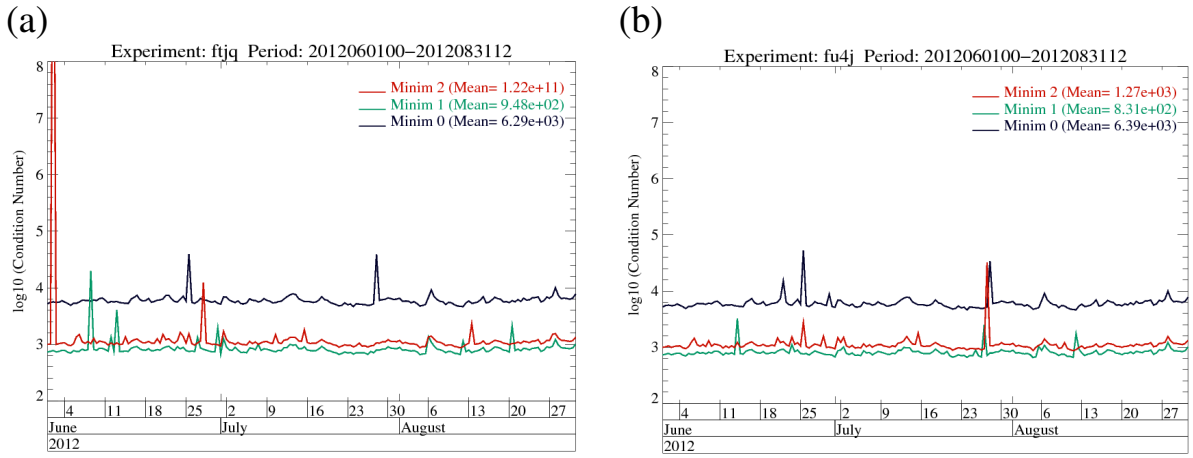


Figure 17: Time series of condition numbers from final T511 L137 38r2 analysis experiments before (a) and after (b) revision of the regularization of exchange coefficients for momentum in the linearized vertical diffusion.

Figure 16a clearly indicates that several CNs were far too large, by up to five orders of magnitude. Fig. 16b displays the new CNs after revision of the regularization of the exchange coefficients for momentum, with LEKPERT set back to TRUE. All previous large condition numbers have now disappeared.

A further verification of the proposed changes was performed with cycle 38r2E (to be used as the baseline for cycle 39r1), which included all modifications to the initial cycle related to physical parameterizations and data assimilation. The corresponding CNs are shown in Figure 17 for June 2012 and confirm the absence of large CN values in the new run (Fig. 17b). It should be noted that the single CN outlier seen in Fig. 17a on 2 June 2012 at 00 UTC is due to a problem with TMI all-sky radiances that is currently under investigation.

The impact on forecast scores is neutral (not shown). The magnitude of the relative changes is usually within 1% and not significant, at all levels and for all parameters over the period June-August 2012. Finally, it was also checked that the revision slightly improves the validity of the tangent-linear approximation, especially at T255 L137 resolution, which is the highest resolution used in the last minimization of the current 4DVAR (not shown).

The proposed revision of the regularization in the linearized vertical diffusion scheme is expected to be implemented with cycle 39r1. It should be noted that no other potential issue could be identified in relation to the use of the linearized physics package with 137 vertical levels.

3.7 Case study: A stratocumulus to cumulus transition case

The impact of the vertical resolution change on the representation of boundary layer clouds was also tested in the Single Column Model (SCM) framework.

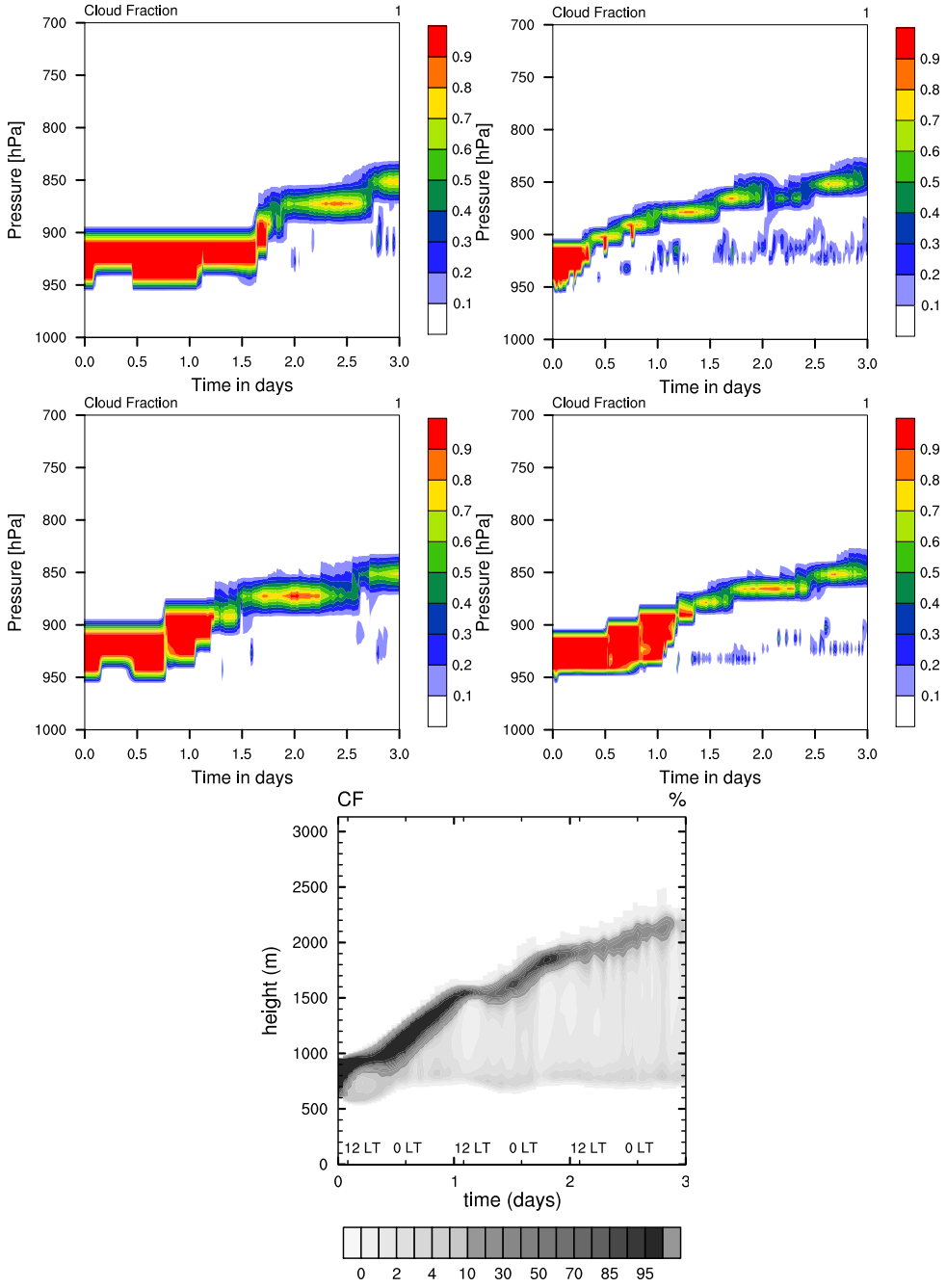


Figure 18: Time evolution of cloud fraction in SCM runs for a stratocumulus-to-cumulus transition case from L91 (top row), left and L137, right - for the control 38r2 SCM; (middle row) L91, left and L137, right - for CY38r2 SCM modified to imbed the changes described in Section 3.2. Bottom panel shows the time evolution of the cloud fraction (%) obtained for this case from a LES.

One of the cases emphasizing best this impact is a stratocumulus-to-cumulus transition case set up in

the framework of the GEWEX Cloud System Studies [17]. Large-Eddy Simulations of this case show a gradual deepening of the boundary layer, associated with the advection of the air mass over warmer ocean waters. The stratocumulus layer gradually thins and eventually breaks up after one day, while shallow cumulus clouds form underneath (Figure 18, bottom). The top panels in Figure 18 show the time evolution of the cloud fraction from the L91 (left) and the control L137 SCM run (right). In the L137 run the cloud evolution is closer to the LES simulation, although the stratocumulus dissipates too fast. This is in agreement with what we had seen in three-dimensional runs where the change in vertical resolution was causing a reduction of cloud cover in marine stratocumulus regions. If the SCM is run with the modified entrainment rate of test parcels used in boundary layer and shallow cumulus schemes (see Section 3.2) the cloud evolution becomes more realistic at both L91 and L137 vertical resolutions (second row in Fig. 18).

Comparing the L137 and L91 runs for this new model version, we can see the benefits of the vertical resolution change for the representation of this type of situation. The inversion is sharper in the L137 run so that the air above the cloud becomes slightly drier and warmer (Figure 19). The entrainment of this warmer and drier air at cloud top favours a more pronounced decoupling of the cloud from the surface layer (see the total water content in Fig. 19). The effect is twofold: the stratocumulus thinning is more marked, and more shallow cumulus clouds pop up at the top of the surface layer.

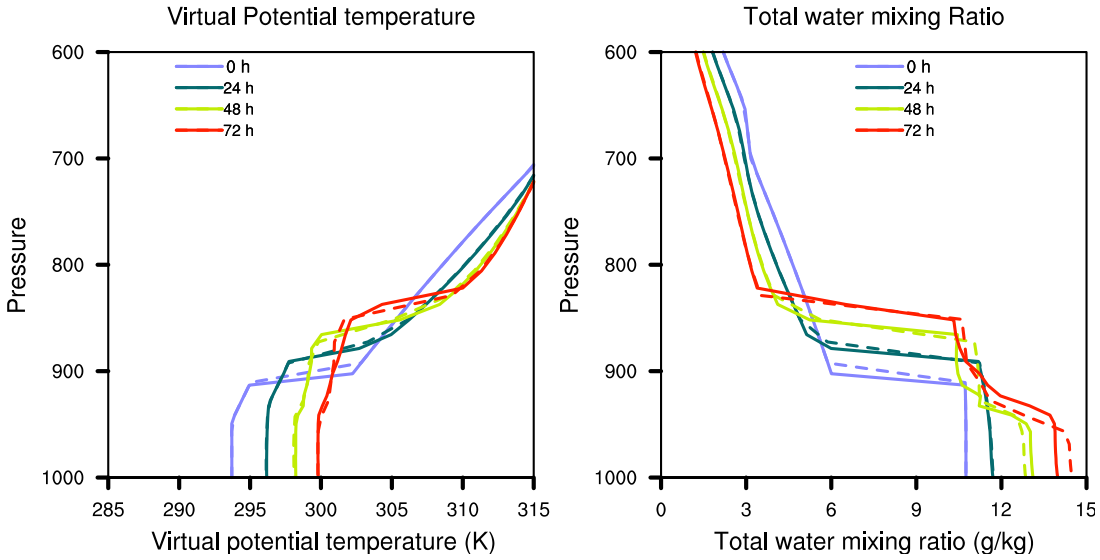


Figure 19: Vertical profiles of virtual potential temperature and total water content in the L91 (dashed) and L137 (full) runs performed with the cycle 38r2 single-column-model, modified to imbed the changes described in Section 3.2, at different time steps (0, 24, 48 and 72 hours).

3.8 Case study: Representation of mixed-phase boundary layer cloud

Supercooled liquid water in boundary layer clouds are abundantly observed in the atmosphere at high latitudes, but the unresolved physical processes remain a challenge to represent in global models. Unresolved processes, such as small-scale turbulence and mixed-phase microphysics maintain the liquid layer at cloud top and must be adequately represented in the model. The presence of liquid directly affects cloud radiative properties in the shortwave and longwave spectrum as well as prolonging cloud lifetimes and is particularly important at high latitudes where radiative impacts of liquid cloud can have a significant effect on temperature near the surface.

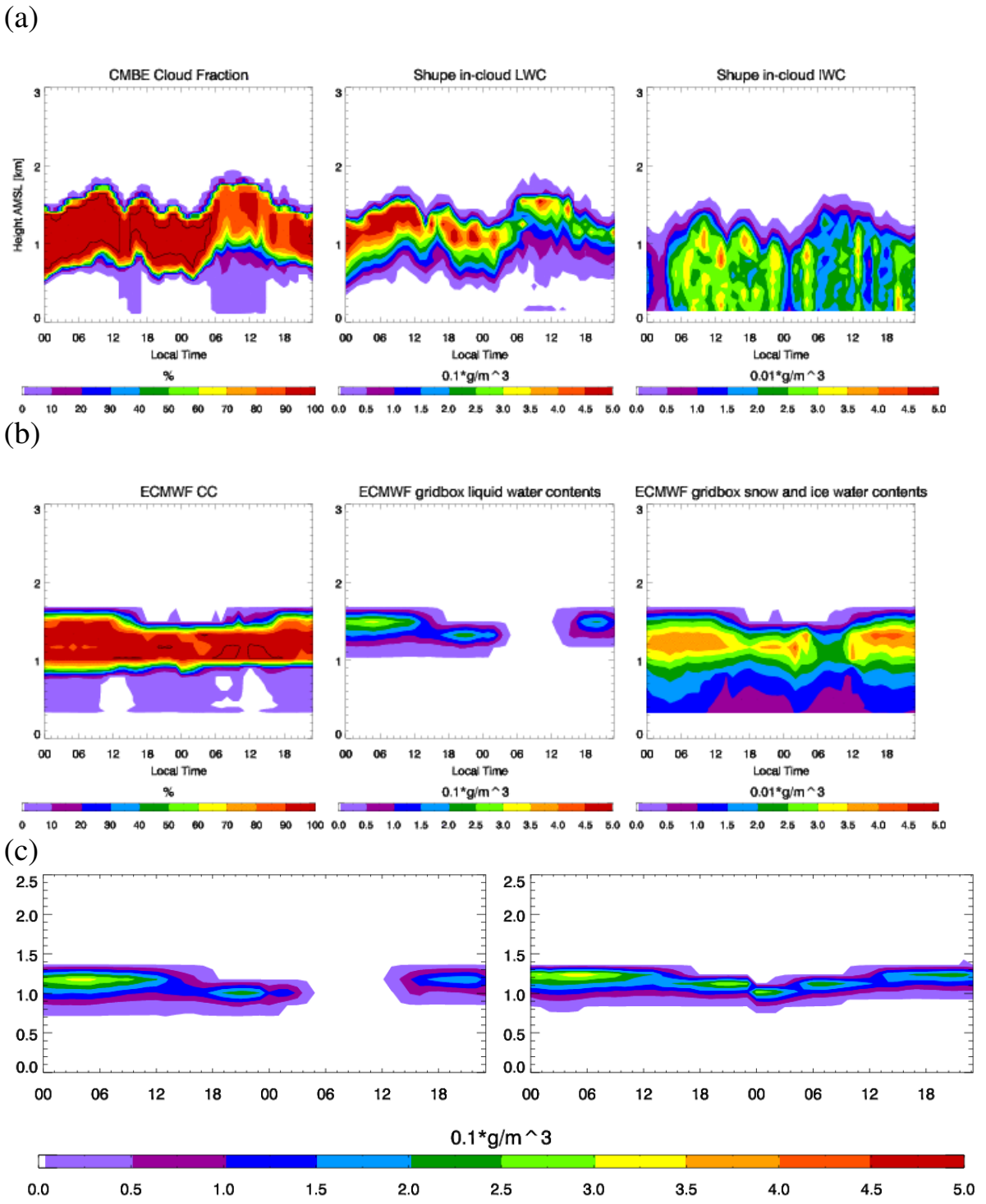


Figure 20: Timeseries of boundary layer cloud profiles during 2 days of the M-PACE observational campaign at the ARM North Slopes of Alaska site during 2004. (a) Derived cloud fraction, super-cooled liquid water and ice water contents from radar and lidar observations, (b) the equivalent timeseries from the 91 level model with reference deposition rate parametrization, (c) a comparison of the supercooled liquid water content timeseries for the model with 91 and 137 levels showing increased cloud-top water contents with higher vertical resolution.

[1] investigated possible reasons why models find it difficult to simulate mixed-phase clouds properly. Low vertical resolution was a dominant factor because nucleated ice particles were not able to fall out of a deep model layer, thus enhancing the ice deposition rate (which depends on the amount of ice present) and reducing the supercooled water. Simulations in a single column model with finer vertical resolution showed increases in super-cooled liquid water at cloud top. We therefore expect that increased vertical resolution in the IFS will also lead to some increase in supercooled liquid water at cloud top.

A case study from the Mixed-Phase Arctic Cloud Experiment (M-PACE) during the autumn of 2004 is used to highlight the impact of vertical resolution in the IFS. A single-layer mixed-phase cloud was observed on October 9/10th 2004, as discussed by [9] for a model intercomparison study. The ECMWF model is initialized at 12 UTC on 9 and 10 October 2004 and integrated for 48 hours from the operational analyses. Forecast lead times from 21.5 to 45.5 hours from each of the two integrations are concatenated to form a two day timeseries.

Figure 20a shows the observed cloud fraction, super-cooled cloud liquid and ice water contents for the two day period derived from radar and lidar. An 800m deep cloud at about -10^{o}C persists through the two days, with a super-cooled liquid water layer at the top of the cloud and ice precipitating out below. The reference model simulation (91 levels, standard deposition parametrization) in Fig. 20b, is able to capture the correct structure of the mixed-phase cloud but has less liquid water and is not able to maintain the liquid water layer for the entire period. However, Figure 20c shows the impact of increased vertical resolution on the super-cooled liquid water layer. In the L137 model, the higher vertical resolution allows increased separation between the precipitating ice and liquid at cloud top, reducing the ice deposition rate and increasing the supercooled water contents in closer agreement with the observations.

4 Data assimilation changes

4.1 Background error covariances

The background error covariance statistics consist of a static part (the correlations) and a dynamic part (the variances) that also represent the flow-dependent component. Recently, the static part has been revised based on model cycle 37r3, which produced a substantial part of the very good impact of cycle 38r1. The flow-dependent component was introduced with cycle 37r2. Both components will become fully flow-dependent with cycle 39r1, later in 2013.

When new covariance statistcs are generated, forecast differences from single integrations following the NMC-method [15] or EDA forecast differences can be used. For the latter, existing background errors are required to run the EDA analyses. These can be taken from an operational version of the background errors unless vertical resolution changes, which was the case here. The generation of the statistics follows 2 steps:

- L137 9-hour forecasts were initialized from L91 EDA analyses at 00 and 12 UTC, sampled every 3.5 days from 20110101 to 20110514. This produces 39 times 9 forecast difference samples by the 10 EDA members for each datum. This sampling is identical to that was used for generating L91 background errors for cycle 38r1. The spectral covariance was calculated off-line and used for initial testing of the assimilation system until L137 wavelet covariances could be calculated. The wavelet covariance version was calculated from the same sample once technical issues with the wavelet generation within IFS were resolved. The resulting background errors were used for all L137 experiments. A global spectral balance operator was derived from the same EDA sample to be used with both covariance matrices.

- The final wavelet covariances were generated from cycle 38r1 L137 EDAs sampled every 1.5 days (for 22 datums from 20110610, 44 datums from 20110910). The L137 EDAs used the L137 background errors derived (see above) from the L91 EDA. The global spectral balance operator was also derived from the same sample.

The idea behind the iterative generation of error covariances is to obtain background errors that are more consistent with the current system and that are initialized using existing stable statistics. However, the L137 covariances based on L137 EDA integrations (i.e. step (2)) produced worse forecasts particularly near the surface (see Figure 21). Note that the L137 EDA analyses were produced with an unmodified model i.e. without the modifications applied to the physical parameterizations as described in this paper. Consequently, only the covariance statistics from step (1) have been used for cycle 38r2. As mentioned above, this cumbersome procedure will become obsolete with cycle 39r1 which introduces fully cycling error covariances.

The structure of the L137 background errors derived from the L91 EDA is, perhaps not surprisingly, very similar to the L91 horizontal and vertical correlations (see Figures 22-23). The L137 global average background error standard deviations are smaller than the L91 standard deviations, Fig. 24, namely 5% smaller for vorticity and unbalanced temperature and about 10% for unbalanced divergence. This is partly due to the fact that the L137 forecasts that are initialized from the interpolated L91 EDA analyses do not include stochastic physics perturbations.

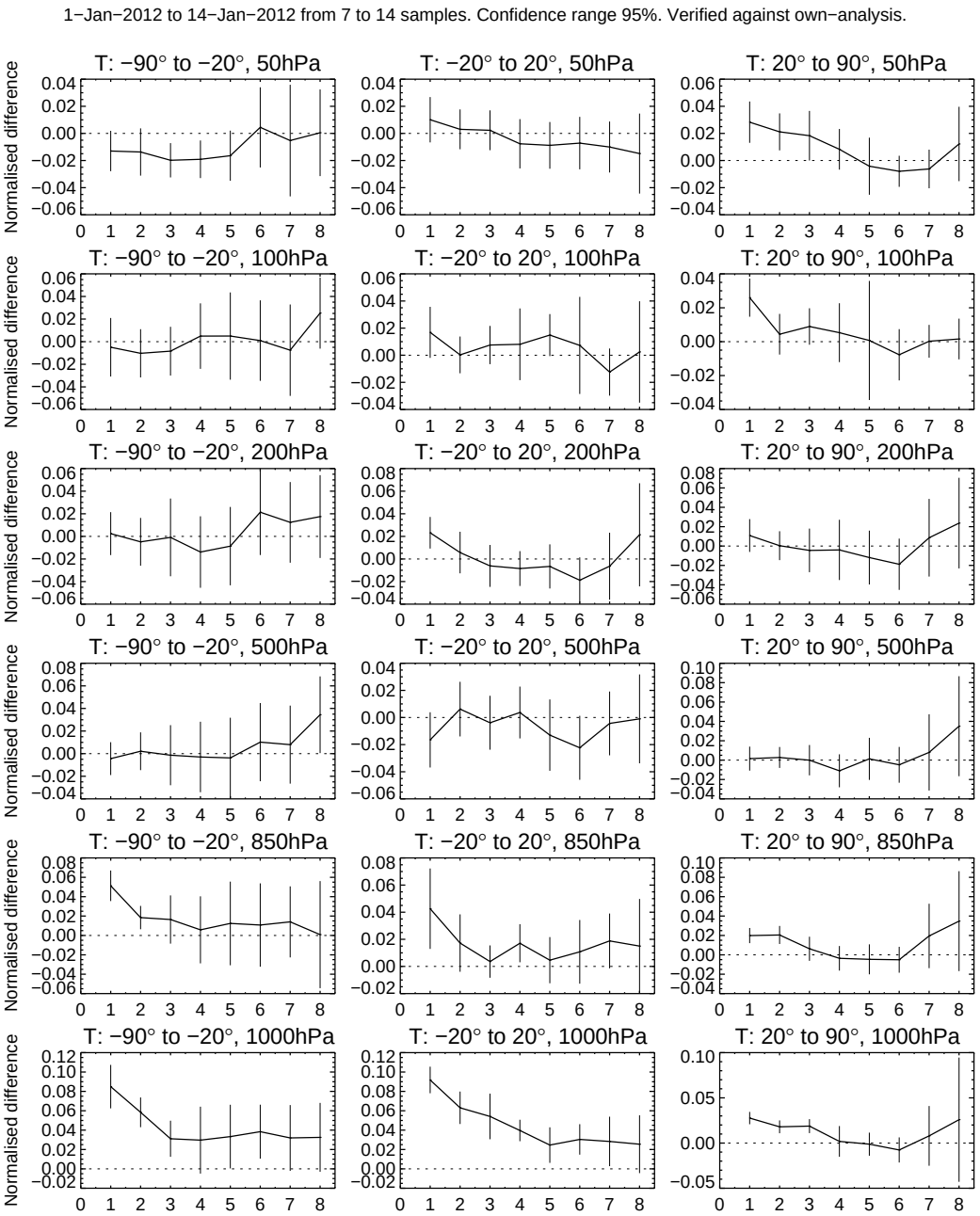


Figure 21: Normalized temperature forecast error differences between new background error covariance matrices derived from L137 and L91 EDA integrations over 15 forecasts.

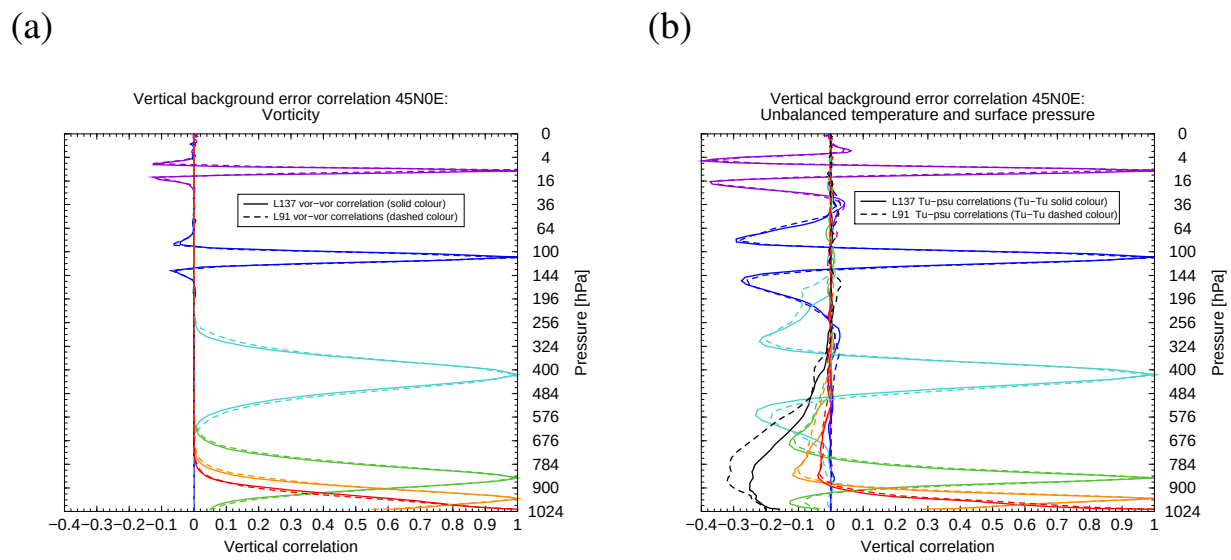


Figure 22: Vertical background error correlations at location 45N/0E for vorticity (a) and unbalanced temperature and surface pressure (b) for the L91 38r1 operational (dashed) and the L137 background error covariances derived from L91 EDA analyses interpolated to L137 for generating L137 9-hour forecast samples (solid).

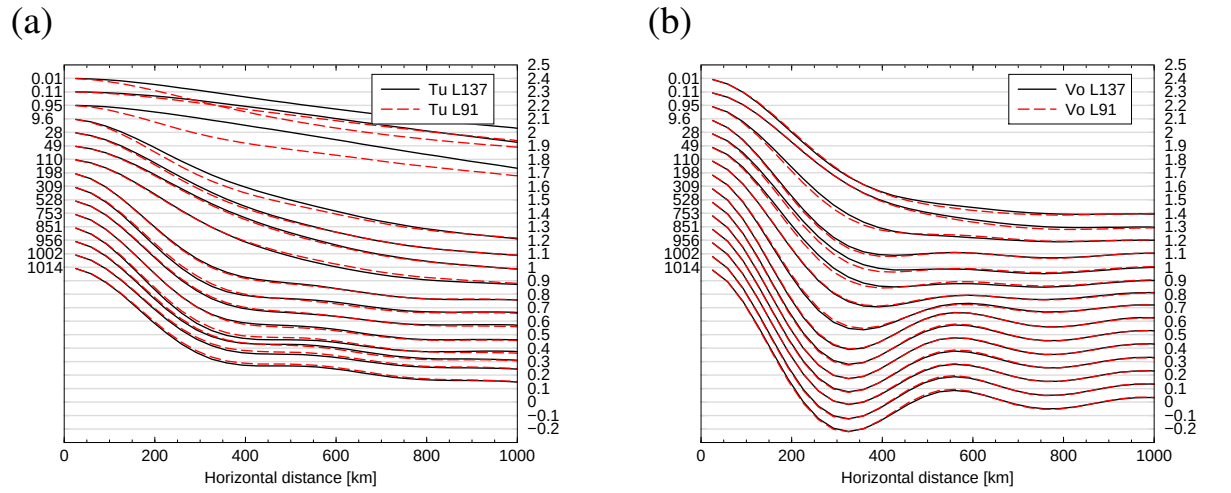


Figure 23: Horizontal background error correlations at location 45N/0E for vorticity (a) and unbalanced temperature and surface pressure (b) for the L91 CY38R1 operational (red dashed) and the L137 (black solid) background error covariances. The correlations are for model levels at 0.01 hPa to 1014 hPa (left axis), with unbalanced surface pressure coinciding with unbalanced temperature at 1014 hPa. For better visibility each correlation is shifted by 0.1 in the vertical, from 0 vertical shift at 1014 hPa to 1.4 shift for 0.01 hPa (right axis).

4.2 Model error cycling in stratosphere

The model error formulation used in the IFS [20] allows for a three dimensional time-dependent forcing of the non-linear model and tangent linear model integrations within the data assimilation window of 4DVAR. The model error is updated each assimilation cycle through the addition of a model error term to the cost-function. It represents a deviation from the assumption of a perfect model in the current data

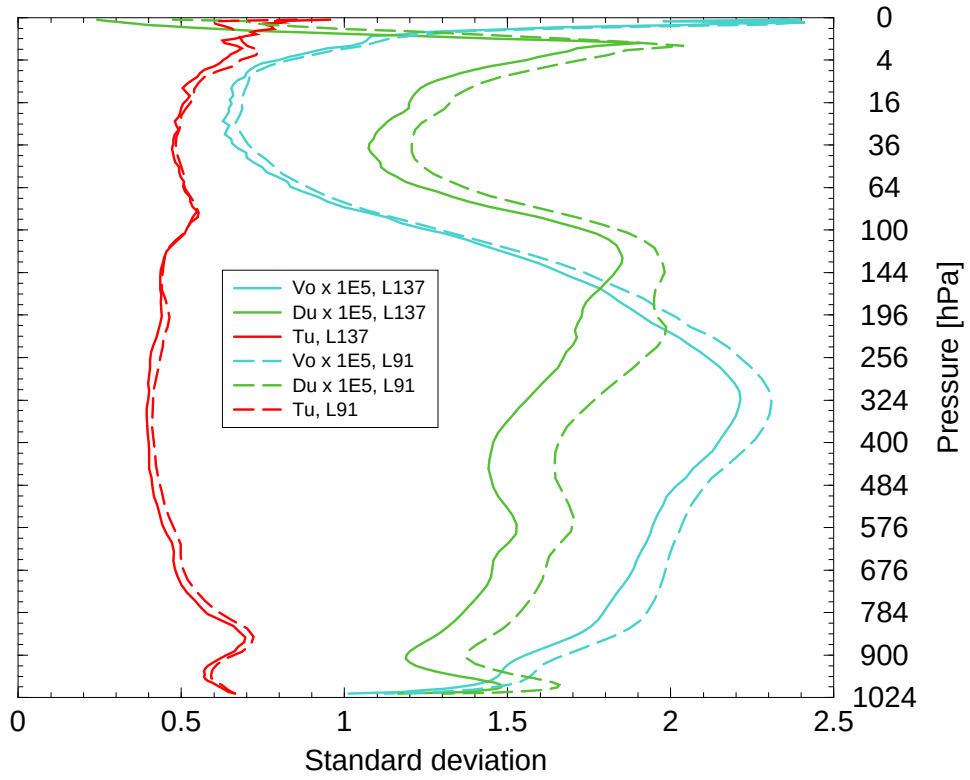


Figure 24: Global average background error standard deviations from the L91 cycle 38r1 operational (dashed) and the L137 (solid) set-up. Units are $10^{-5}s^{-1}$ for vorticity, $10^{-5}s^{-1}$ for unbalanced divergence, and K for unbalanced temperature.

assimilation system and is essential for extending the length of the 4DVAR window.

The current implementation of the model error is a simplification of the more general framework in that it is kept constant over the length of the assimilation window and that it is only active in the stratosphere above 40 hPa. The value of the forcing is updated each cycle to provide a first-guess value for the next cycle. The assumption is that it mainly captures systematic errors.

The first L137 assimilation experiments showed a worse first-guess fit to observations above 50 hPa for radiosonde wind and for AMSU-A channels. After investigating several possible causes, it was found that turning the model error off restored the first-guess fit quality.

At this point, experiments to identify the cause or solution for why the L137 model error degrades performance have not been successful. Turning the model error off for cycle 38r2 thus appeared to be a safe option, in particular as the current model errors already have a small impact on the performance of the L91 system.

Further investigations will be required to reinstate the model error and the current results have given some suggestions for further research. The model error covariances at L137 have maximum standard deviations at low wavenumbers 2-10, compared with wavenumbers 20-100 at L91. Due to how the model error covariances are designed (subtracting forecast tendencies several days into the forecast),

this shift most likely reflects a change in the dominant term of the model error at L137. Comparing the semi-diurnal tide in the L91 and L137 (without model errors), the L137 analysis has 10%-20% larger maximum amplitude of the semi-diurnal tide. One hypothesis is that the L91 model has about the same amplitude of the semi-diurnal tide as observations, whereas the L137 model has a larger amplitude than the observations. During the analysis, the model error attempts to correct the tide, as indicated by its semi-diurnal tide-like structure. However, even if the model error gets the spatial structure correct at the start of the assimilation window, it is currently held constant throughout the window and can thus not capture the time evolution of the semi-diurnal tide. The fit of the model to the semi-diurnal tidal component of the observations is thus not improved, and a positive feedback with growing model error and analysis increments takes place

4.3 Initial data assimilation experiments

After the generation of new background error covariance statistics the L137 system was found to have less skill than the L91 system. At this stage it was not clear if this originated from the analysis system or the forecast model. To answer this question, cross-forecasts were performed, i.e. in addition to L91/L91 and L137/L137 analysis/forecast experiments, runs with L91 forecasts initialized by L137 analyses and L137 forecasts with L91 analyses were performed (see Table 2).

Figure 25 show the results as normalized temperature forecast error differences with respect to the L91 analysis/L91 forecast experiment. Negative values mean that the experiment perform better than the reference. The most apparent feature is that those experiments that employ a L137 model perform very similarly at all levels and latitude bands and they produce larger temperature errors at lower levels. The only exception is seen at 200 hPa in the Tropics. The L91 forecasts that were initialized with L91 analyses show a smaller deterioration at lower levels until day-3. This performance difference is not seen for geopotential or wind, except at 1000 hPa in the Tropics. It is thus mostly the result of the small change in temperature mean errors that resulted from the resolution upgrade as described in Section 2.1 and that has been mostly recovered by model changes as described in Section 3.

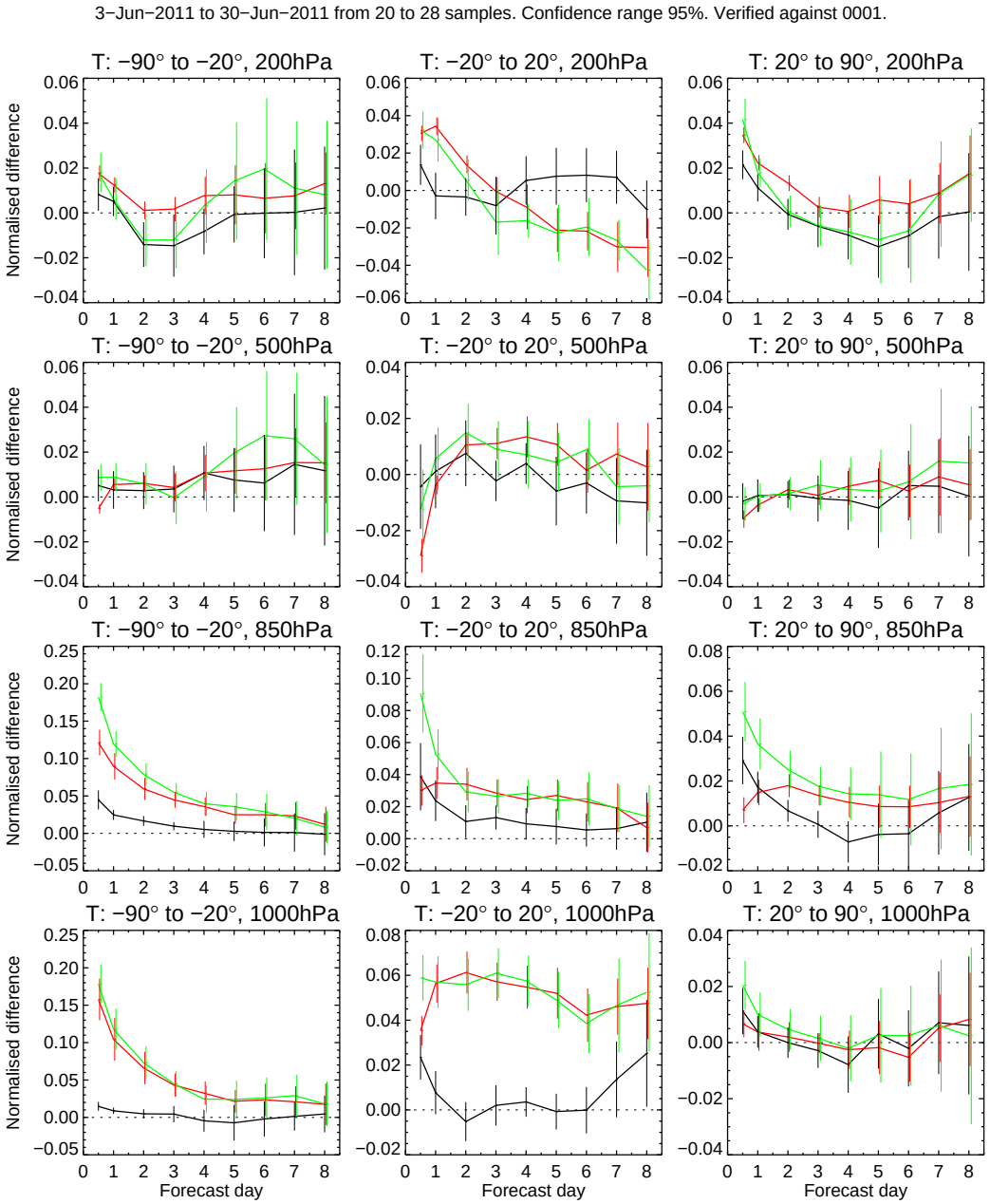


Figure 25: Normalized temperature forecast error difference from cross-experimentation. All errors are relative to the L91 analysis/L91 forecast experiment, namely L137 analysis/L91 forecast (black), L91 analysis/L137 forecast (red) and L137 analysis/L137 forecast (green). Time period is 20110603-20110630, vertical bars indicate 95% statistical significance.

5 Final assessment

Since the introduction of the EDA, full-resolution e-suites have become more computing intensive than before because 4DVAR assimilation and medium-range forecast suites require running an EDA ahead that provides background error variances of the day. When substantial model changes are implemented also the EDA needs to be rerun in order to (i) produce error covariance statistics that represent the current model cycle and (ii) filter and calibrate the background error estimates to compensate for sampling noise and systematic errors. Both aspects were particularly important for the vertical resolution upgrade. In the future, activating the explicit feedback of errors and calibration between 4DVAR and EDA will alleviate this issue. However, for cycle 38r2 a separate L137 4DVAR experiment needed to be run ahead of the L137 EDA to allow the calibration statistics to be computed. The L137 EDA then produced the errors of the day for another 4DVAR experiment that represented the actual e-suite experiments (see all experiment identifiers in Table 2). This system has been applied for January-March 2012 and June-December 2012 in RD, followed by the OD e-suite for January-June 2013.

The high-resolution experimentation has been complemented by 3-member ensemble climate runs, verified with ERA-Interim, over the 30-year period 1981-2010, by 128 forecasts of the 51-member ensemble through boreal winter, summer 2012 and winter 2013, as well as lower-resolution monthly forecasts, namely 45-day 15-member ensemble integrations initialized in February, May, August and November, respectively, for the period 1989-2008.

During experimentation it was found that verification quality was affected by whether the verification was performed on pressure or model levels. This became particularly evident near sharp gradients, near the surface and where orography shows large variability. In these cases the interpolation software can create artefacts that can be misinterpreted as model performance issues. The topic of pressure vs model-level verification will be dealt with in a separate paper.

5.1 Data assimilation experiments

5.1.1 38r2 with L91 vs 38r1 with L91

The scorecard in Figures 46-48 (see Appendix) shows the performance of the model changes when both cycle 38r2 and 38r1 are run with 91 levels. This is an important test since the ENS and reanalyses are likely to be upgraded to L91, and SEAS may maintain this resolution as well. It was therefore important to establish that the modifications implemented with 38r2, which aim at improving the L137 model, do not deteriorate the performance once run with 91 levels. The verification has been performed with observations to exclude the issue of model vs analysis set-up.

The evaluation confirms that the performance between the two cycles is very similar with L91. Cycle 38r2 performs slightly better in the lower stratosphere in all areas, which is most likely the combined effect of the radiation and non-orographic gravity wave drag revisions. Both act mostly on the mean state and thus the mean and the root-mean-square error (RMSE) are most improved.

5.1.2 38r2 with L137 vs 38r1 with L91

In general, the results from the HRES experiments suggest that cycle 38r2 is superior to 38r1. Figure 26 shows the zonal-mean change in day-1 forecast RMSE for temperature, zonal wind, vertical velocity, and geopotential for 20120101-20120321. In general, errors are reduced for all these quantities. Note

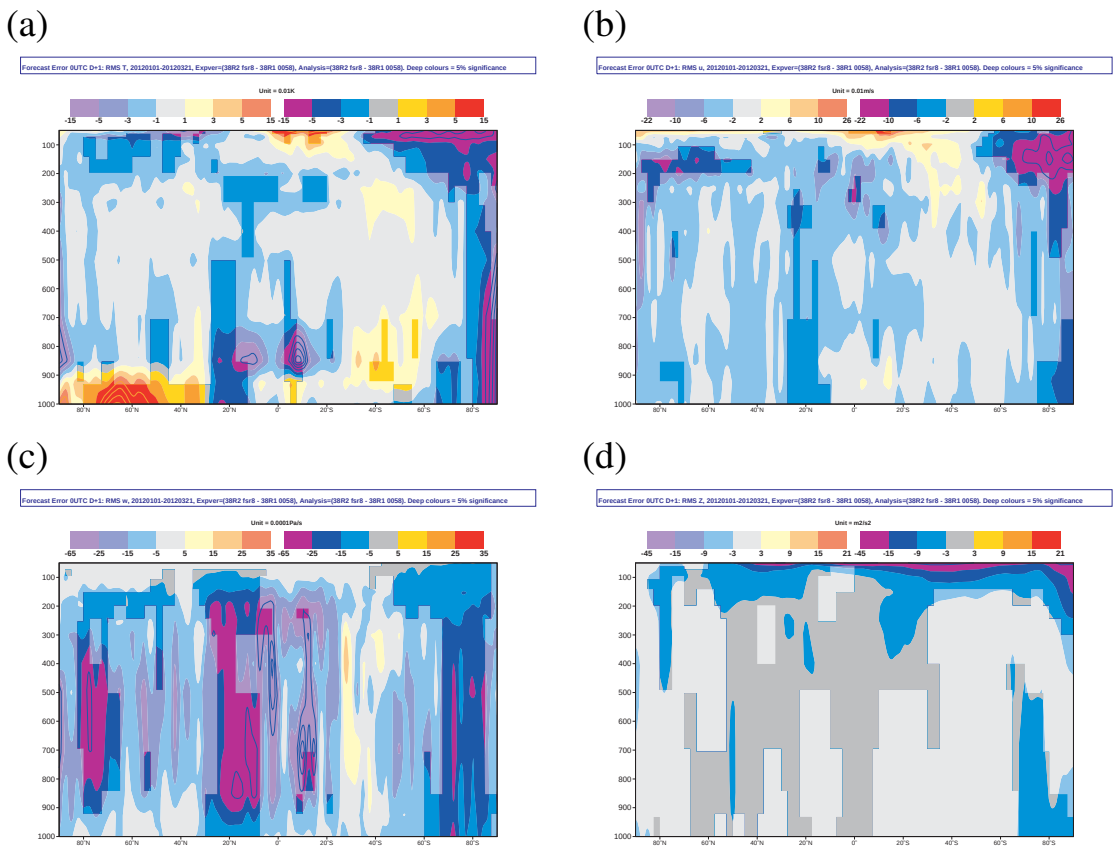


Figure 26: Change in zonal mean day-1 forecast RMSE (38r2-38r1) for (a) temperature, (b) zonal wind, (c) vertical velocity, and (d) geopotential. Results are based on all high-resolution forecasts initialized at 0 UTC over the period 20120101-20120321.

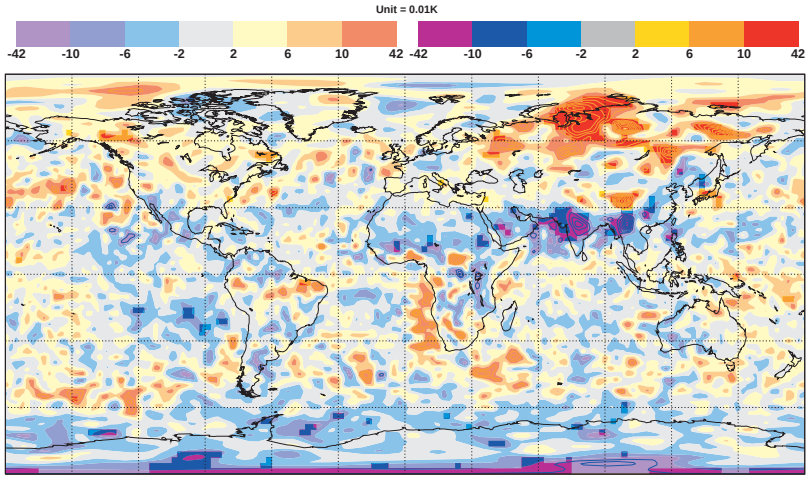


Figure 27: Change in day-1 forecast RMSE (38r2-38r1) for temperature at 925 hPa over the period 20120101-20120321.

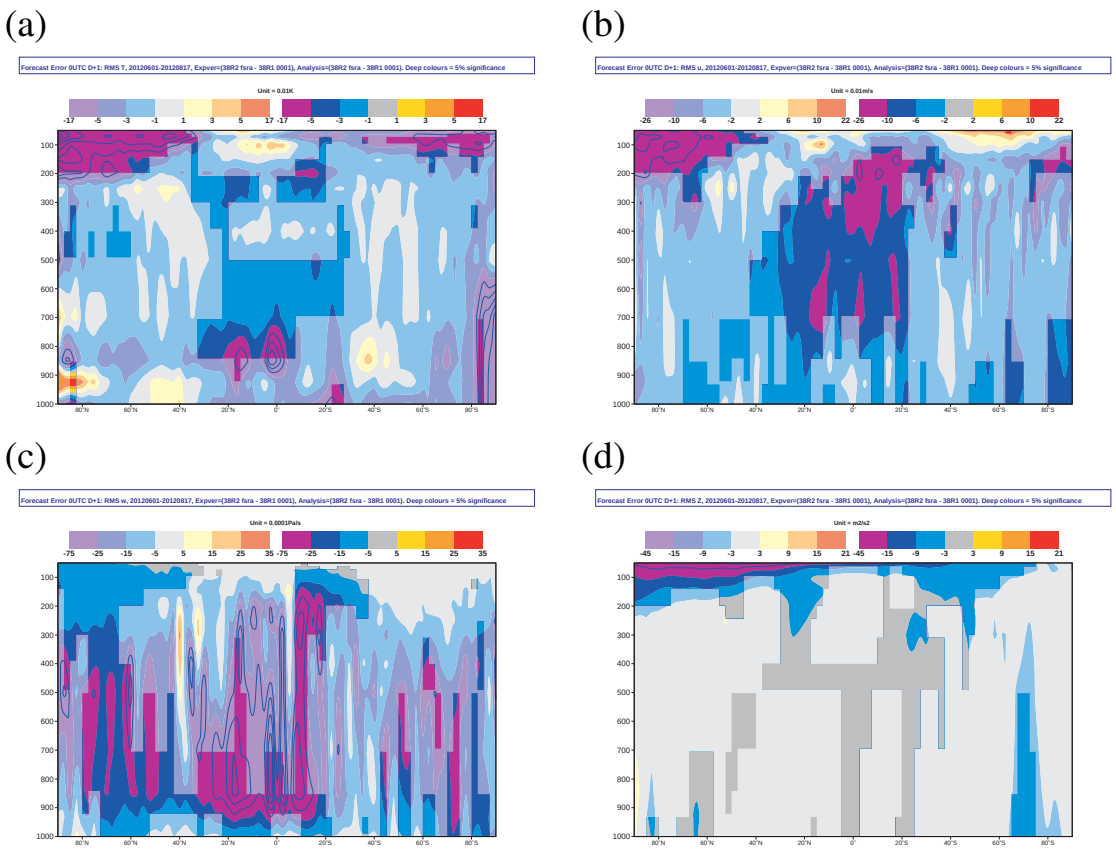


Figure 28: As Fig. 26 for the period 20120601-20120817.

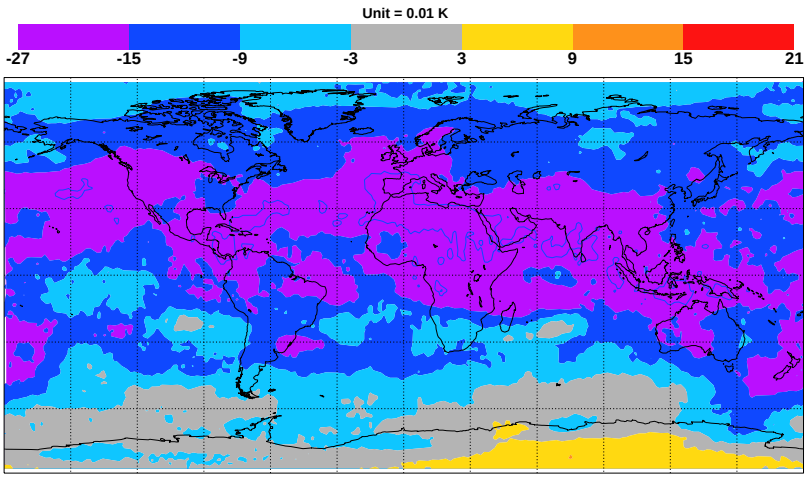


Figure 29: Change in day-2 forecast RMSE (38r2-38r1) for AMSU-A channel 9 (which is a measure of temperature over the layer 30 to 150 hPa, approximately) over the period 20120701-20120731.

that bold colours indicate statistical significance at the 5% level. The change in meridional wind error at day-1 (not shown) is very similar to the change in zonal wind error. The main issue would appear to be an increase in RMSE for lower-tropospheric temperature at around 60N. Figure 27 suggests that this

issue is confined to Siberia. Since surface pressure in the Siberian High is higher than the pressure levels where the problem is manifested, the issue is not simply due to potential changes in extrapolation below the surface. Further investigation of this issue would be useful.

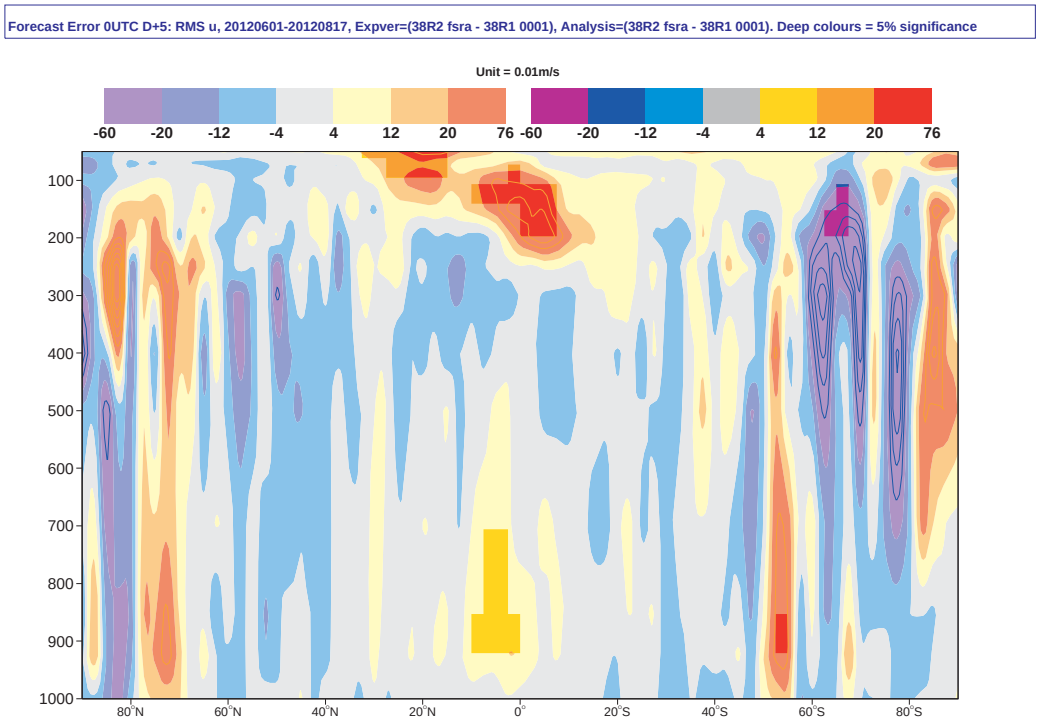


Figure 30: Change in zonal mean day-5 forecast RMSE (38r2-38r1) for zonal wind over the period 20120601-20120817.

The corresponding changes in zonal-mean RMSE for the summer season, 20120601-20120817 (Figure 28) are, if anything, better than for the period 20120101-20120321. One possible issue (although it is not statistically significant in the model cycle) is an increase in RMSE for temperature around 100-150 hPa. However, Figure 29, which shows RMSE relative to AMSU-A channel 9 microwave observations (most sensitive to temperatures integrated over the layer 30 to 150 hPa, approximately), indicates a reduction in error. It is possible that the increased error relative to the model analysis is therefore associated with a sharper and thus more realistic representation of the tropopause with the increased number of levels.

At longer lead-times, differences in zonal-mean error become statistically less significant for the available data. However, one issue that does emerge is a statistically significant increase in zonal mean RMSE of the zonal wind between about 100 and 200 hPa in the tropics (Figure 30 shows the results for the period 20120601-20120817, but the other season shows a similar signal). Figure 31 (solid lines) shows how this increased error emerges with increasing lead-time for the 200 hPa streamfunction. The red dots indicate that the difference is statistically significant for all lead-times between day-4 and day-10. A possible explanation is offered by the dotted forecast 'activity' curves. As lead-time increases, forecast error should asymptote to the activity. Since the model with higher vertical resolution (38r2) is more active, there will be a tendency for cycle 38r2 to have larger errors relative to 38r1 as lead-time increases.

While in Fig. 31 the forecasts tend to show more activity than the analyses (dashed curves), this latter result is not repeated for the other season and is presumably associated with the decrease in inherent

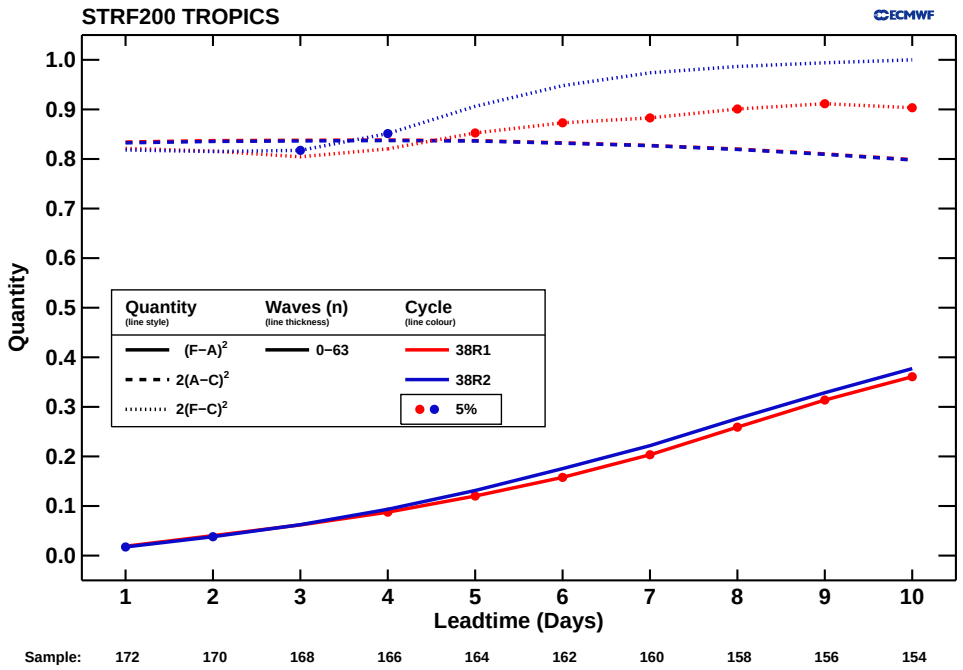


Figure 31: Mean-squared error (solid) and mean-squared anomalies from climate (i.e. 'activity': dashed for analysis, dotted for forecast) for cycle 38r2 (blue) and 38r1 (red) for 200 hPa streamfunction in the tropics over the period 20120601-20120827.

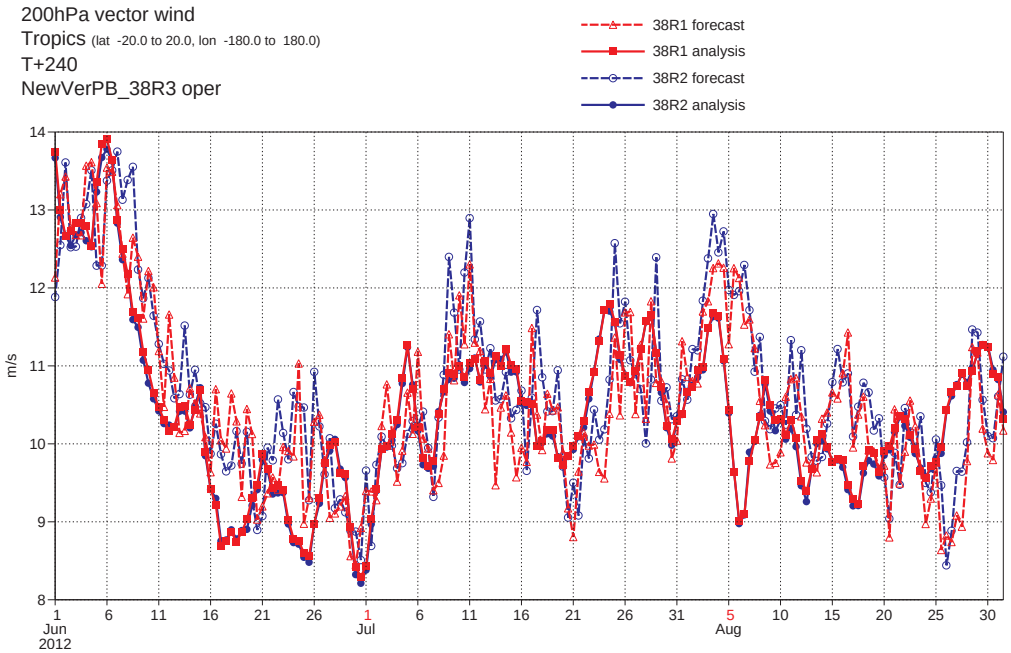


Figure 32: Analysis (solid) and day-10 forecast (dashed) activity of 200 hPa vector wind in the tropics for cycle 38r2 (blue) and 38r1 (red) over the period June-August 2012.

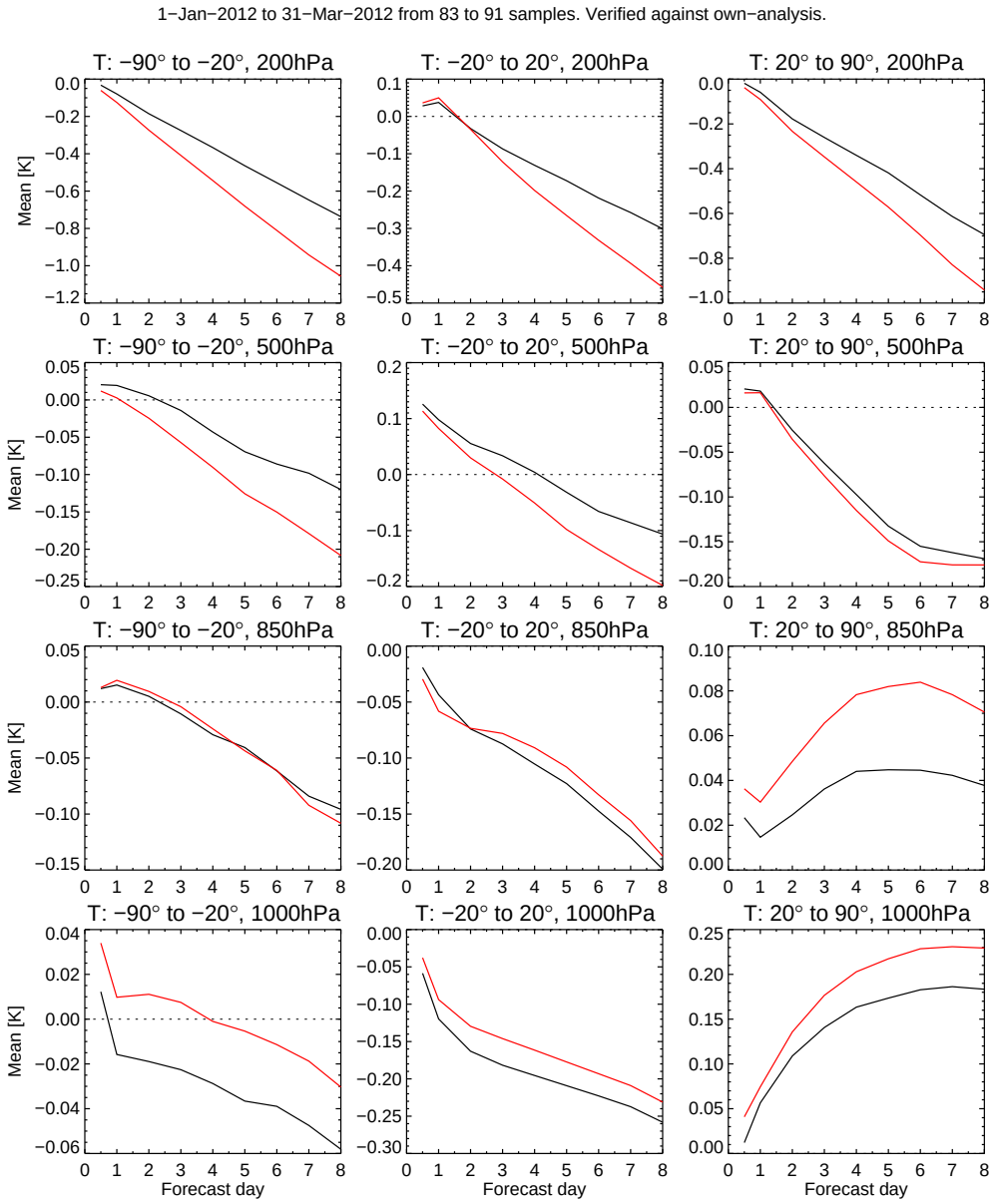


Figure 33: Mean temperature forecast error for cycles 38r1 (red) and 38r2 (black) for 200, 500, 850, 1000 hPa (top-to-bottom) and southern hemisphere, tropics, northern hemisphere (left-to-right), period: January-March 2012.

predictability of each seasonal-mean anomaly with lead-time. Looking at the activity of winds (Fig. 32) themselves shows that, indeed, forecast activity has increased slightly between 38r1 and 38r2 (by about 2% in the medium range) and that day-10 forecast activity is about a few percent higher than the analysis activity throughout June-August 2012. However, activity is generally very consistent between analysis and forecasts for temperature, geopotential and wind in the ECMWF system (not shown) and this is only marginally changed by the new cycle.

Figures 49-52 (see Appendix) summarize the overall performance of high-resolution experimentation for cycle 38r2 compared to 38r1 when verified with own analyses (Figs. 49-50) and observations (Figs. 51-

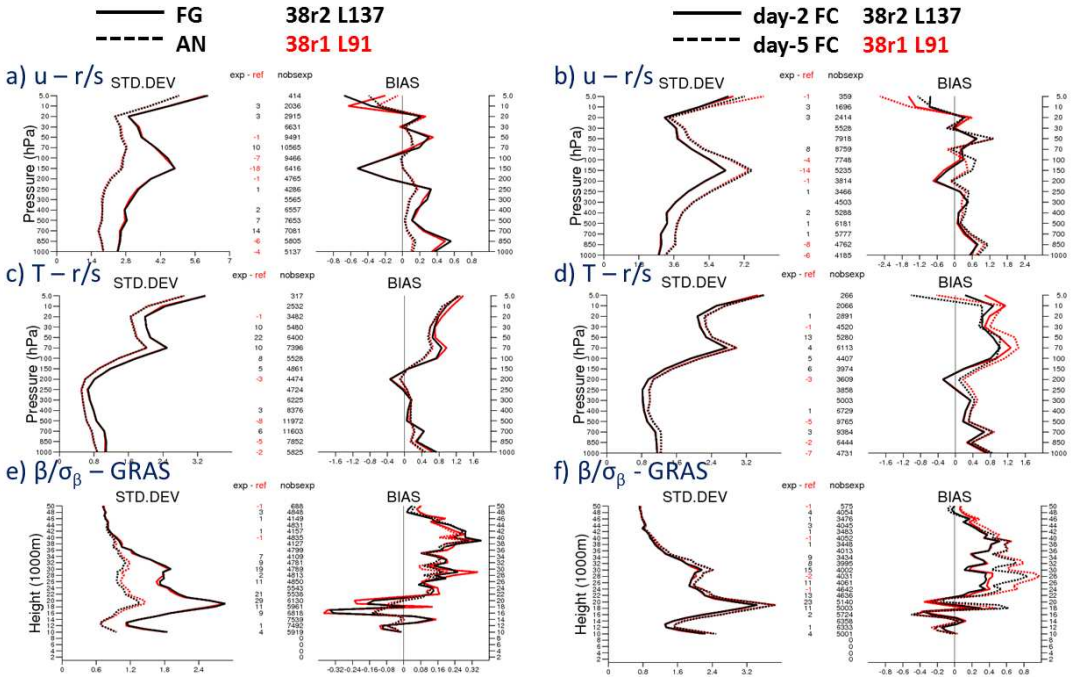


Figure 34: Evaluation of model analysis and first-guess (i.e. short-range forecast) (a, c, e) as well as 2 and 5-day forecasts with observations (b, d, f) for radiosonde u-wind component (a, b), radiosonde temperatures (c, d) and normalized radio occultation bending angles (e, f). Period January-March 2012 in the Tropics for 38r1 (red) and 38r2 (black).

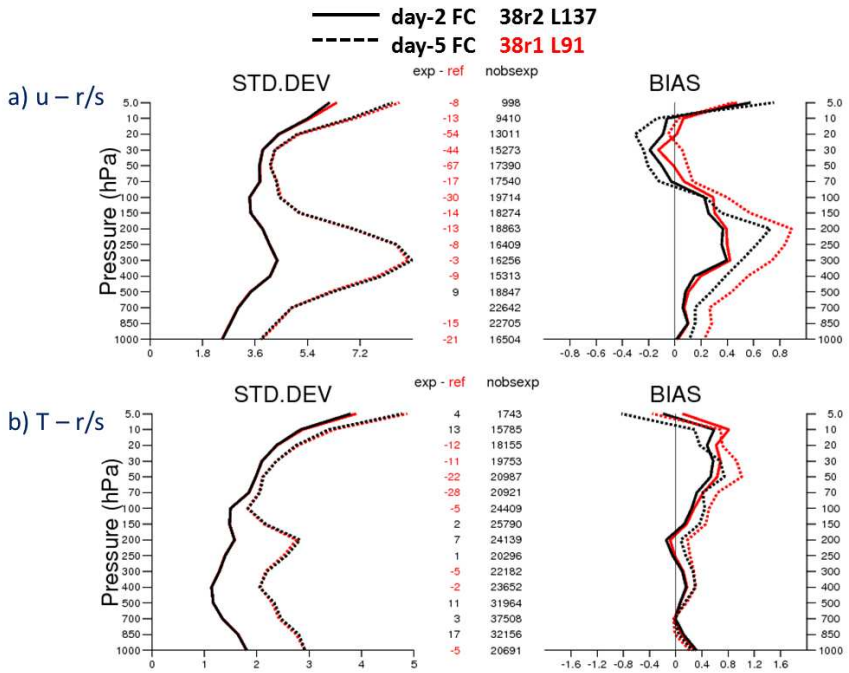


Figure 35: Evaluation of 2 and 5-day forecasts with observations for radiosonde u-wind component (a), radiosonde temperatures (b). Period January-March 2012 in northern hemisphere for 38r1 (red) and 38r2 (black).

52), i.e. radiosonde data. The use of radiosonde data for verification implies a sampling bias since most radiosonde observations are available over Northern hemisphere land areas and are dominated by datasets from the US, Europe, East Asia, South Africa, Australia etc.

Figs. 49-50 can be summarized such that in the extra-tropics small deteriorations are found for upper level humidity and lower level temperatures over the southern hemisphere, and improvements are obtained for temperature, wind and geopotential across all areas and most levels. The impact lasts 2-5 days. The strongest relative skill difference is seen in the tropics, which reflects the zone of primary impact of model resolution on tropospheric stability as described in Section 2. Here, deteriorations are found for upper level moisture and temperature as well as lower level winds. This is partly contributed by changes in mean error as well as increased variability as indicated by the change of error standard deviations.

For temperatures, the change in mean errors is shown in Figure 33 at 4 pressure levels and southern hemishere, tropics, and northern hemisphere. The lower level cooling in cycle 38r2 amplifies negative errors and causes a deterioration with respect to cycle 38r1 while the warming at upper levels reduces errors. In the mid-troposphere the effect is fairly neutral. This mechanism applies similarly to all latitudes but the contribution of the mean error to the rms-error decides whether this is visible in the rms-error differences between the cycles. Note that across the UTLS, the forecast error changes sign close to analysis time (too warm upper troposphere, too cold lower stratosphere) that evolves into a generally too cold UTLS into the forecast. This change of sign also drives the sign of impact of cycle 38r2 relative to 38r1.

Verification against observations (Figs. 51-52) exhibits a generally more neutral performance. The skill changes only little at lower and mid levels and improvements for temperature and geopotential are found at all latitudes. The only obvious deterioration here is for temperature at lower levels in the Tropics. There is an apparent inconsistency between the verification with analyses and observations for temperature and wind at lower levels in the Tropics that may be partly explained by the spatial sampling of radiosonde verification.

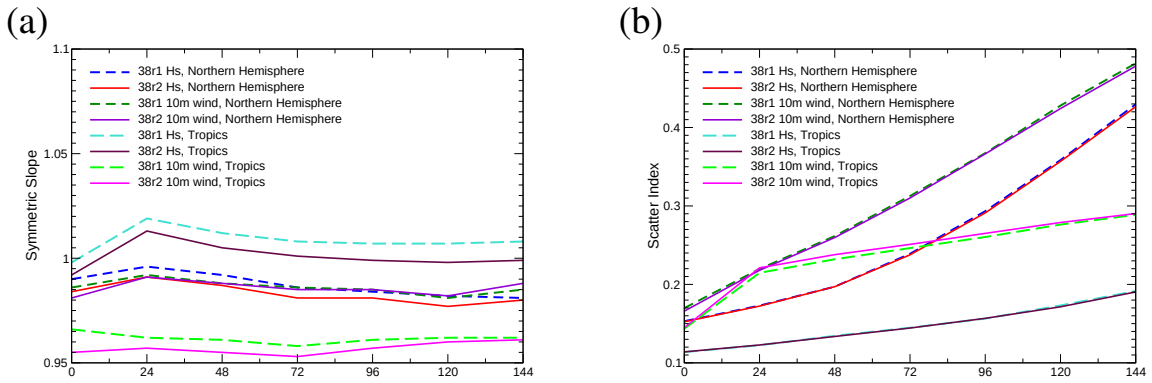


Figure 36: Verification of 10-metre winds and significant wave heights with buoy observations for January-March 2012 and June-August 2012.

Figure 36 shows verification of 10-metre winds and significant wave heights (Hs) for January-March 2012 and June-August 2012 between cycles 38r1 and 38r2. The scatter index is a normalised standard deviation of error whereas the symmetric slope is a measure of the activity (the square-root of the ratio of the model variance to the observed variance). Overall, a small improvement was found with the new cycle except for tropical winds. The latter does not depend on ocean basin.

To further investigate the difference between standard verification with analysis and observations, Fig-

ure 34 shows vertical profiles of observational verification, as it is applied in the data assimilation system, for day-2 and day-5 forecasts for selected parameters and observing systems in the Tropics. In the analysis, wind error standard deviations change very little between model cycles and a small deterioration of 850 hPa winds is seen (Fig. 34a). However, the errors are of the order of 0.5 m/s and the deterioration is less than 0.05 m/s. Into the forecast (Fig. 34b), biases increase to 1 m/s but the difference between the cycles remains at 0.05 m/s and does not grow with forecast range. For temperatures (Fig. 34c, d), a more apparent improvement is seen, confirming Fig. 51. This improvement becomes more pronounced into the forecast range and it extends lower into the troposphere. This is also confirmed by statistics based on Global Positioning System (GPS) radio occultation observations that are very sensitive to atmospheric temperature gradients in the upper troposphere and stratosphere. Figure 35 confirms that this improvement is also seen at higher latitudes and that it is a rather stable feature of the new model cycle.

5.1.3 EDA estimation of the unbalanced control vector in 4DVAR

A long standing objective of data assimilation at ECMWF has been to provide flow-dependent estimates of background errors to the 4DVAR analysis. This is important for many reasons, such as the quality control of the observations and because flow-dependent errors help determine the correct weight to give to the observations presented to the analysis. The goal of using flow-dependent errors had been partially achieved in cycles 37r2 and 38r1 with the use of spatially filtered and rescaled sample standard deviation fields from the ECMWF Ensemble of Data Assimilations (EDA) [6, 3].

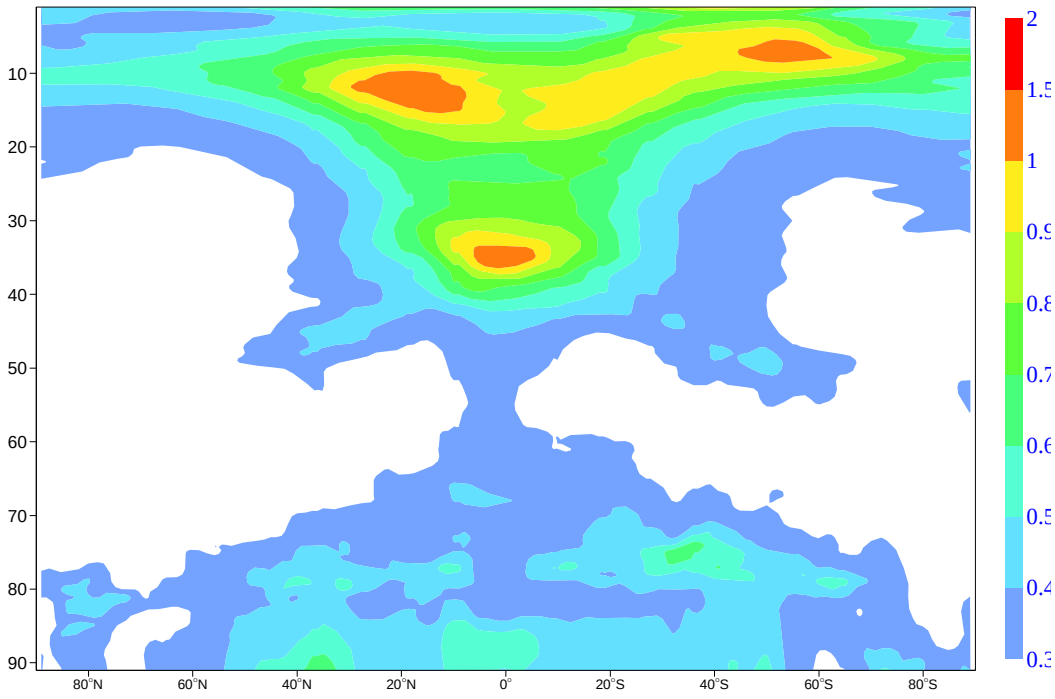


Figure 37: Zonal average of unbalanced temperature errors estimated from the EDA as a function of model level. Values are averaged over May 2012. Units are degrees K.

However a major limitation of those implementations resided in the fact that only the balanced part of the 4DVAR control vector (i.e., the vorticity field) was estimated online from the EDA. The unbalanced part of the 4DVAR control vector (i.e., unbalanced divergence, D_u, unbalanced temperature, T_u, and

unbalanced surface pressure, SP_u) was modelled by spatially homogeneous climatological estimates. This limitation has been removed in cycle 38r2, where online estimates of D_u, T_u, SP_u from the EDA are also used.

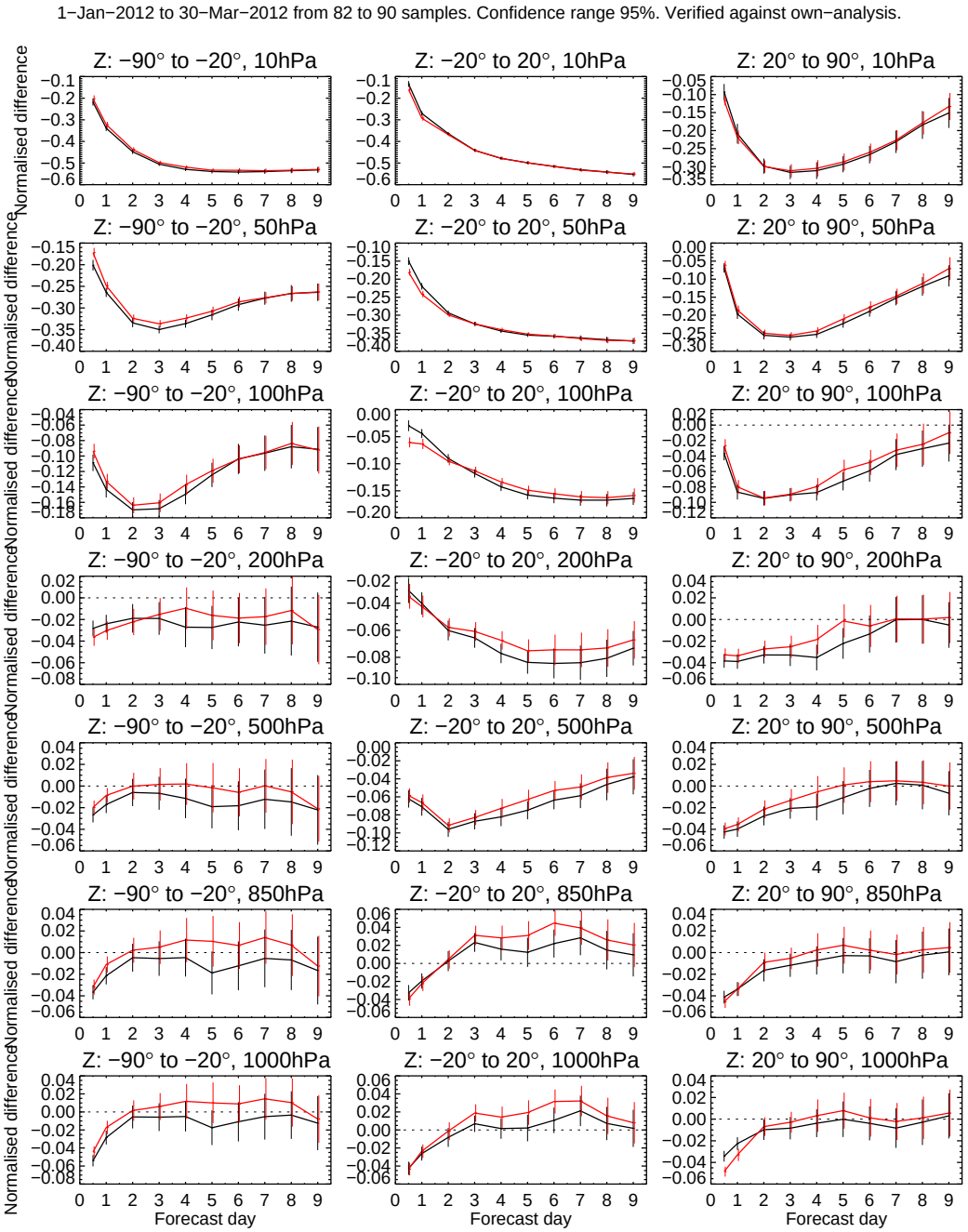


Figure 38: Normalised root mean square error reduction of geopotential height forecasts of cycle 38r2 experiments over the 01/01/2012 - 31/03/2012 period. Black curves refer to the experiment using EDA unbalanced errors, red curves to the experiment not using them. Reference is the ECMWF cycle 38r1 pre-operational suite. Error bars represent 95% confidence levels. Verification against own analysis.

It is important here to highlight two aspects. In Figure 37 we present the longitudinal average of un-

balanced temperature errors estimated from the EDA as a function of model level, averaged over the month of May 2012. Excluding the higher stratosphere (levels > 20), the largest impact is visible in the boundary layer and in the tropical lower stratosphere. This is consistent with the expected breakdown of the mass wind balance relation in the boundary layer and the signature of convection generated gravity waves in the tropical stratosphere.

The impact of EDA unbalanced errors on the IFS analysis and forecast skill has been tested in isolation from all other changes introduced in this cycle because they are the most important meteorological contribution to cycle 38r2 not directly connected to the change in vertical levels. To this end two sets of experiments for cycle 38r2 have been run at full operational resolution (T1279L137): one with the EDA unbalanced errors active, the other with these errors switched off. In Figure 38 we show the geopotential height normalized forecast error reduction scores for the two esuites (black curves refer to the experiment with EDA unbalanced errors, red curves to the experiment not using them) over the January-March 2012 period. The reference experiment is the pre-operational cycle 38r1 suite available for the period in question. The positive impact of the EDA unbalanced errors is clearly visible throughout the troposphere and lower stratosphere.

5.2 Model climate on the seasonal time scale

In order to evaluate the impact of the new cycle on the model climate, seasonal experiments have been run, both in uncoupled and coupled mode. The data set is based on 3 ensemble members and forecasts are initialised twice a year for 30 years (1981-2010). For more information about the experimental setup see [13].

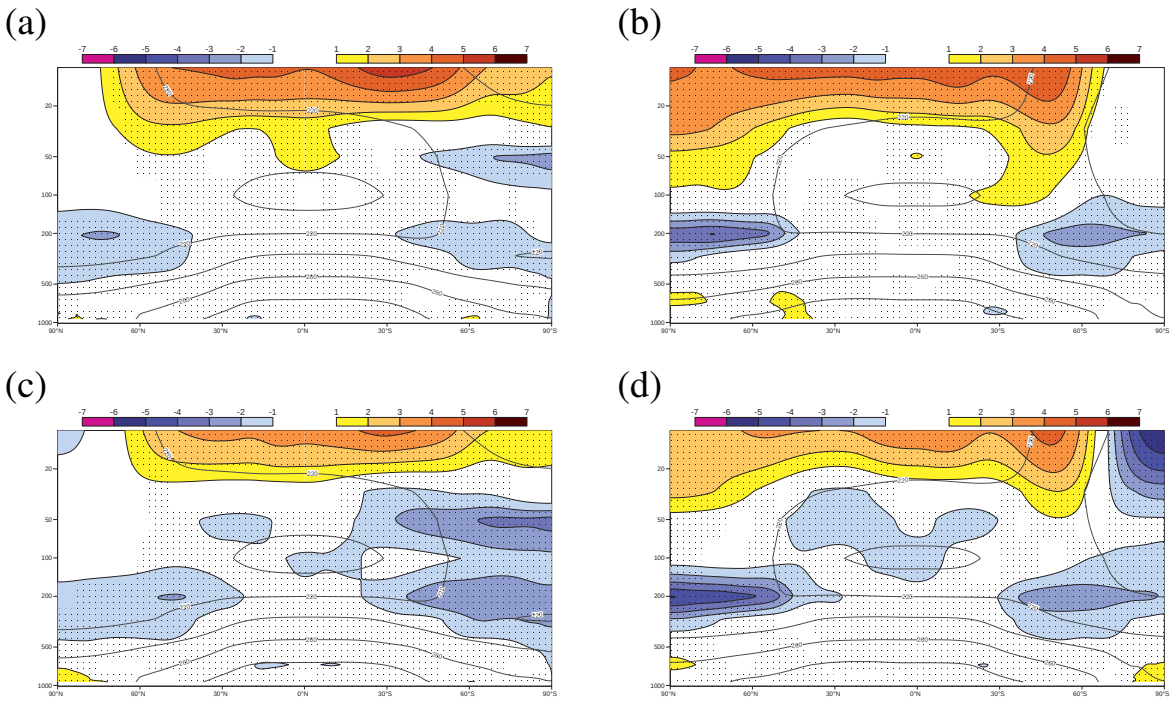


Figure 39: Zonal cross-section of temperature bias, verified with ERA-Interim climatology (1981-2010), with cycle 38r2 for December-February (a) and June-August (b), and with cycle 38r1 for December-February (c) and June-August (d). x-axis: latitude, y-axis: pressure in hPa, dots: 5% statistical significance.

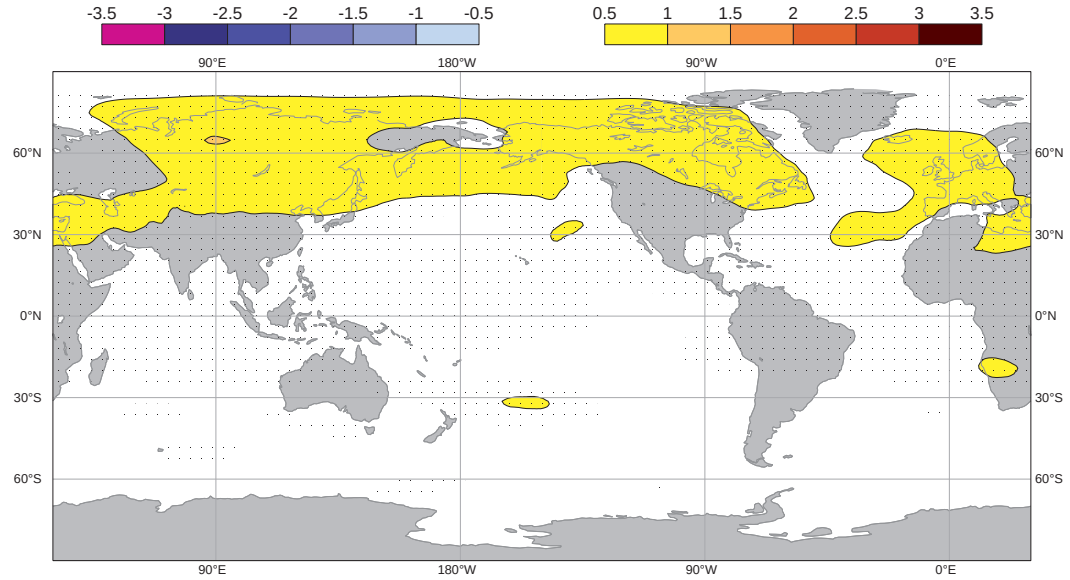


Figure 40: Climatological 500 hPa temperature difference between cycles 38r2 and 38r1 for June-August 1981-2010. Dots: 5% statistical significance.

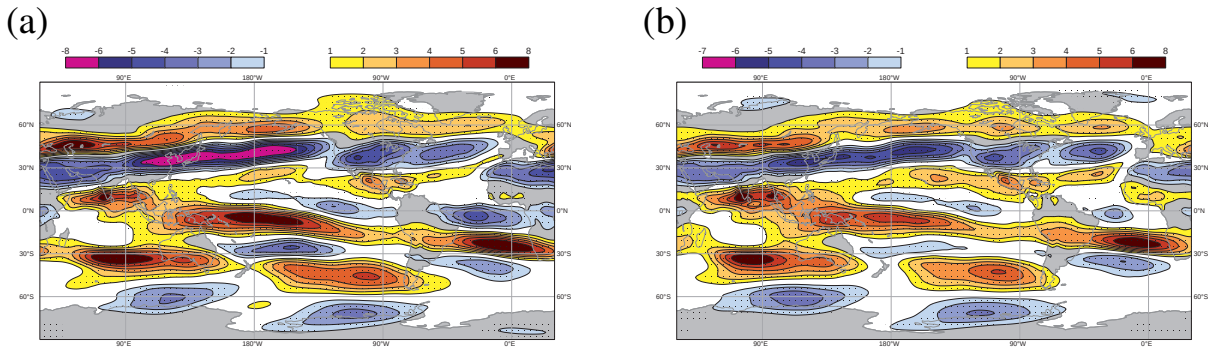


Figure 41: Climatological 200 hPa zonal wind bias, verified with ERA-Interim climatology (1981-2010), for cycles 38r1 (a) and 38r2 (b). Dots: 5% statistical significance.

Figure 39 shows a vertical cross-section of the zonally averaged temperature bias with respect to ERA Interim for the coupled model integrations for cycyle 38r1 (Fig. 39a, b) and 38r2 (Fig. 39c, d), for December-February (DJF) and June-August (JJA), respectively. The most pronounced biases are around the tropopause in the extra-tropics and a warm bias in the upper stratosphere. Comparing the two cycles, 38r2 has improved stratospheric temperature errors during polar nights (north pole in DJF, south pole in JJA), especially over the south pole. Another improvement is the reduced cold bias in the lower stratosphere in the tropics. In the upper stratosphere the exisiting warm bias has increased but there are uncertainties associated with the ERA-Interim dataset due to a lower model top (0.1 hPa) and limited data availability constraining the analysis at these altitudes. All presented changes are also seen in the uncoupled integrations or integrations with L91 for cycle 38r2. The latter means that the change in bias is not caused by the vertical level upgrade but by the accompanying modification of non-orographic gravity wave drag (see Section 3.4).

A negative change in the model climate is a small enhancement of the positive temperature bias in the troposphere in the northern hemispheric summer north of 40N. The difference between the two model

cycles for 500 hPa temperature is shown in Figure 40 (the difference is also present in the uncoupled model and in the 91 level runs). Linked to this feature is an increased bias in the position of the Northern hemisphere jet stream in summer, as seen in the plots for zonal bias in Figure 41. The change is most pronounced over Asia and the Pacific. This difference could be related to an enhanced zonal circulation in the stratosphere (e.g. at 50 hPa, not shown). Another negative change in 38r2 is a slightly increased easterly (negative) wind bias in the lower troposphere over the tropical western Pacific; a change that also is present in the short-to-medium range.

5.3 Medium-range ensemble forecasts

Model cycle 38r2 was tested extensively with the ensemble prediction system. This section describes the evaluation of forecast lead times up to 15 days. The following section describes the impact on the monthly forecasts.

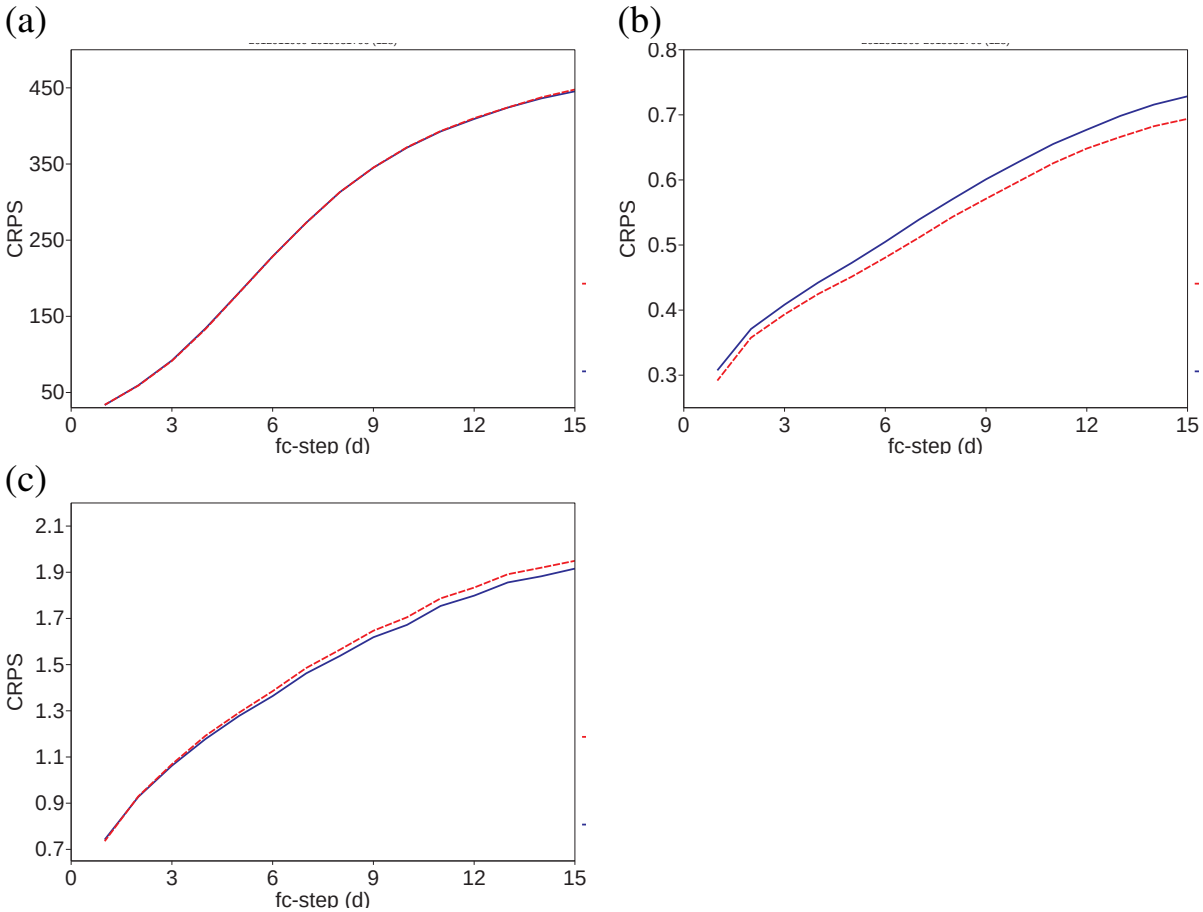


Figure 42: Continuous ranked probability score for (a) 500 hPa geopotential in the northern extratropics, (b) 850 hPa temperature in the tropics and (c) 850 hPa zonal wind in the tropics. Combined sample of 128 cases; cycle 38r2 (red) and 38r1 (blue).

The vertical discretization of ENS remained at 62 levels in cycle 38r2. An increase of the number of levels is planned for the cycle following 38r2. The ENS experiments have been initialised with analyses and EDA perturbations from the respective model cycles, these have 137 levels for the 38r2 experiment and 91 levels for the 38r1 experiment. The ENS experiments have a resolution of T639/319; they consist

of 50 perturbed members and one unperturbed control forecast.

Cycle 38r2 contains two minor changes of the ENS configuration. Firstly, the formulation of vertical mixing was changed in the tangent-linear and adjoint models used in the singular vector computation. The simplified vertical diffusion scheme of [4] has been replaced by the vertical diffusion scheme that is also used in 4DVAR [14, 8]. Secondly, the scaling factor that determines the amplitude of the singular vector initial perturbations has been increased by 10% in order to compensate for a reduction of the analysis error standard deviation estimate provided by 4DVAR. This results in singular vector initial perturbation of similar amplitude in cycles 38r2 and 38r1.

Experiments with the ENS cover three periods: 10 January – 14 March 2012 (boreal winter 2012), 20 June – 13 August 2012 (boreal summer 2012) and 10 January – 17 March 2013 (boreal winter 2013). All forecasts start at 0 UTC. Start dates are 48 hours apart in the two periods in 2012 and 24 hours apart in the 2013 period. The corresponding experiment identifiers are listed in Table 2. No major differences have been identified between the three different periods. Here, results for the combined sample of 128 cases are summarized.

The spread of the ensemble is very similar in the two model cycles. Overall, cycle 38r2 leads to small reductions of ensemble spread except for temperature at 850 hPa (not shown). The magnitude of the relative changes in spread remains below 1.5%.

The impact of cycle 38r2 on probabilistic skill is close to neutral in the northern extra-tropics. Figure 42a shows the neutral result for the continuous ranked probability score (CRPS) of 500 hPa geopotential for the northern extra-tropics (20–90N). Winds at 200 hPa show statistically significant minor improvements during the first 3 days and minor degradations during the last 2 days of the range. In the southern extra-tropics, the impact is close to neutral overall except for temperature at 850 hPa, which is degraded at all lead times. In the tropics, the signal is mixed with a considerable positive impact on temperature at 850 hPa (Fig. 42b) and a considerable negative impact on winds at 850 hPa (Fig. 42c).

5.4 Monthly forecasts

The impact of cycle 38r2 on the monthly forecast skill scores has been evaluated from a 15-member ensemble experiment starting on the 1 February, 1 May, 1 August and 1 November 1989-2008, respectively. The integrations are 45 days long with a T399 horizontal resolution for the first 10 days of integration and T255 after day 10. The model has been integrated with 62 vertical levels. Probabilistic skill scores have been computed for weekly means (day 5-11, 12-18, 19-25 and 26-32) from an experiment using cycle 38r2 and an experiment using cycle 38r1. Ranked probabilistic skill scores (RPSS) and Relative Operating Characteristics (ROC) scores of most variables are generally very close in both experiments, suggesting a neutral impact of 38r2 on the monthly forecast skill scores (see Figure 43). Cycle 38r2 has a slight positive impact on 500 hPa geopotential height skill scores for day 5-11 and day 12-18. The most noticeable impact of 38r2 is for zonal winds at 50 hPa in the Tropics, with an improvement for 38r2 in all 4 weekly periods.

The impact of cycle 38r2 on the skill of the model to predict the Madden-Julian-Oscillation (MJO) has also been assessed using the same set of re-forecasts. Results suggest a neutral impact of cycle 38r2 on the MJO skill scores (RMS error, bivariate correlation) and the amplitude of the MJO.

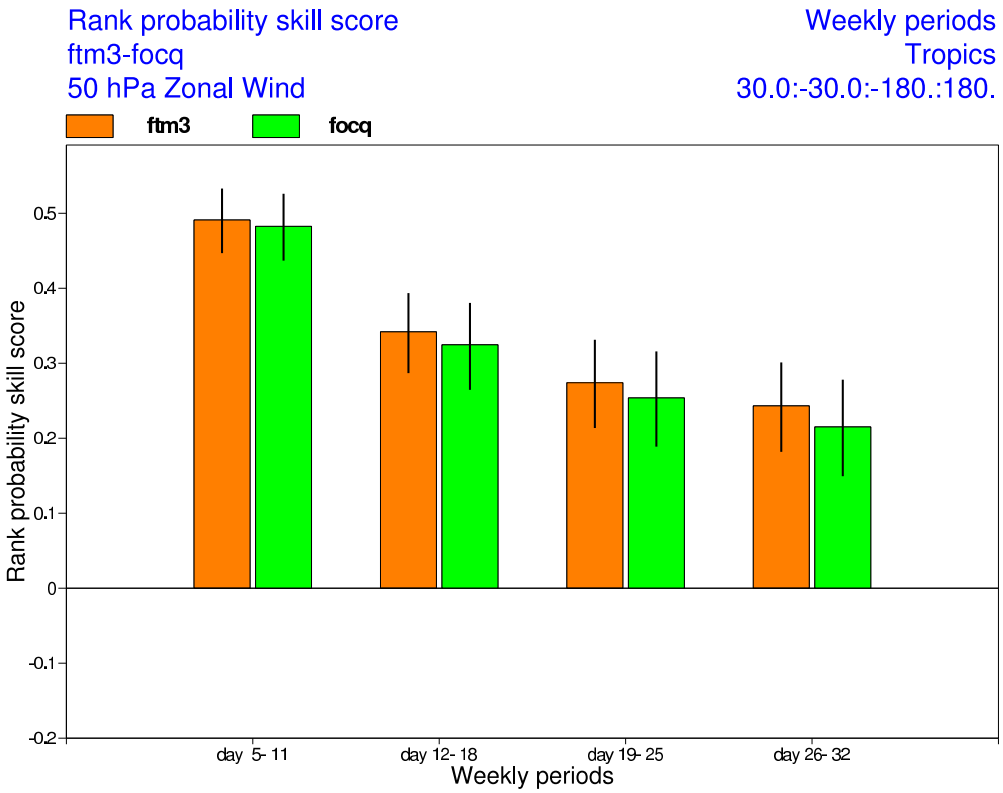


Figure 43: Ranked probability skill score of 50 hPa zonal wind over the Tropics. Orange bars shows cycle 38r2, green bars show cycle 38r1. The vertical black lines represent 2 standard deviations.

5.5 Additional diagnostics

5.5.1 Tropical cyclones

Figure 44 shows the accumulated error statistics from 13 months of high-resolution tropical cyclone forecast evaluation against observations.

A small improvement of position error in the medium range is noted for the new cycle while cyclone core pressure errors remain unchanged. Major changes of the tropical cyclone prediction skill were not expected since the model's activity remained largely unchanged and none of the changes affect, for example, latent heat release. The impact of the changes applied to background error covariances and EDA were unlikely to influence smaller scales.

5.5.2 2-metre temperature, January 2011

To address the question whether the new cycle affects the prediction of 2-metre temperatures over areas with stable boundary layers and persistent low-level mixed-phase clouds that had caused large negative night-time biases in January 2011 over Northern Scandinavia, dedicated experiments at lower resolution have been run. Figure 45 shows the analysis verification with model cycle 37r2 that was operational at the time, cycle 37r3 that included a modification of supersaturation and ice deposition rate in clouds,

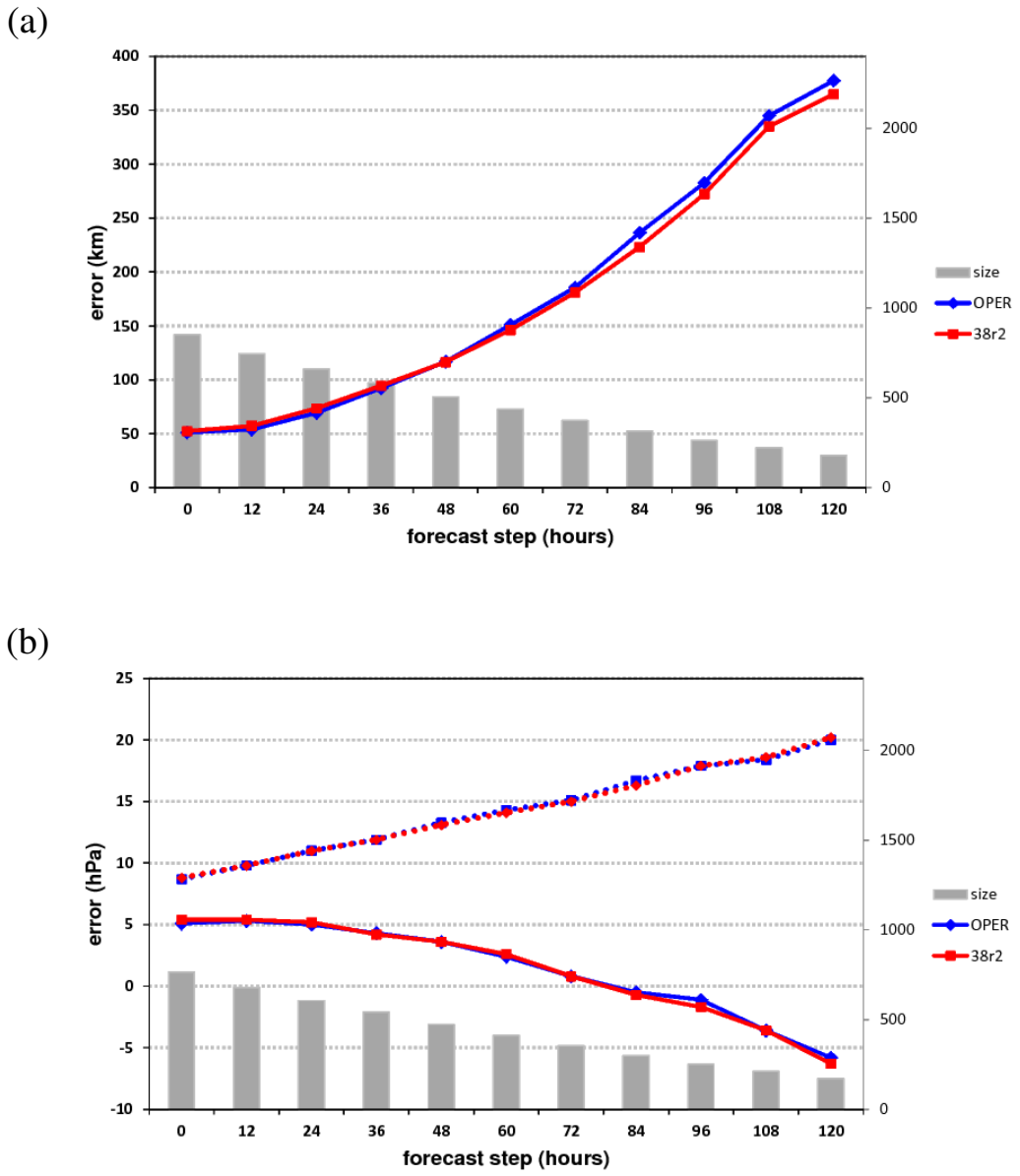


Figure 44: Mean tropical cyclone centre position (a) and central pressure error (b), obtained from RD and OD experimentation with cycle 38r2 (red) compared to operations (cycle 38r1 = OPER; blue) as a function of forecast range. Sample size is indicated by grey bars.

as well as cycles 38r1 and 38r2. The latter has been run with 137 levels. The results indicate that the performance of cycle 37r3 has been retained in this situation and that forecast biases remain almost unchanged across the last three model versions. If cycle 38r2 is run with 91 levels a similar performance is obtained (not shown).

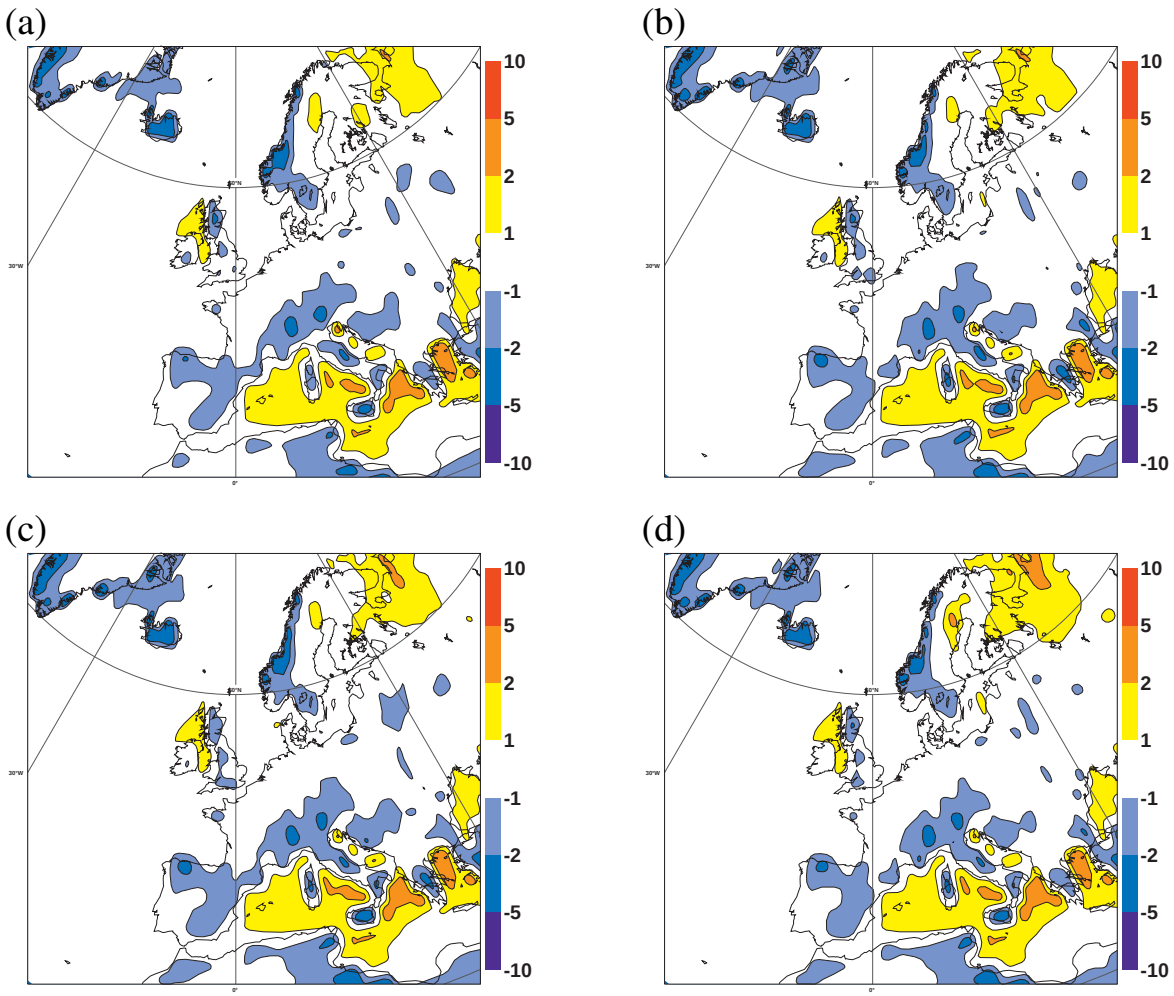


Figure 45: Mean 48-hour 2-metre temperature forecast errors valid at 00 UTC from T511L91 experiments with cycle 37r2 (a), 37r3 (b), 38r1 (c) and a T511L137 experiment with 38r2 (d) over Europe in January 2011.

6 Summary and Outlook

Vertical resolution upgrades are not trivial to perform without deteriorating the system when measured with standard skill metrics. Small changes in mean state (model climate) and activity can penalize root-mean-error and anomaly correlation scores and there is still insufficient global observational verification available in areas and at levels where vertical resolution upgrades make a difference.

The new ECMWF model cycle 38r2 has been implemented oprationally on 25 June 2013 after about 1 year of dedicated research and testing. The initial results mostly exhibited small vertical temperature state differences that were most obvious in the tropics and that also affected wind, shallow clouds and thus radiation. Modification of shallow cumulus, boundary layer and convection parameterizations proved to mostly compensate for these differences. Elimination of vertical level dependent functionality in sub grid-scale orography and radiation schemes as well as an enhanced non orographic wave drag formulation added further benefits. The examples of enhanced skill for marine boundary layer and mixed phase clouds demonstrate the potential of this vertical resolution upgrade for a better representation of physical processes in the near future.

The inclusion of unbalanced EDA errors further improved cycle 38r2 with 137 levels over 38r1 with 91 levels in most areas, except for lower stratospheric and tropical boundary layer winds. Unbalanced EDA errors are not necessary for 137 levels, but were included now to prepare for near future observation and background error developments. All other aspects of the background error covariance statistics were updated using new 137 level EDA forecasts, including new static correlations and tuning of dynamic variances. Changes to background errors are more complicated when model vertical resolution changes and the final configuration (which also included turning off model errors in the data assimilation) had to be chosen from competing alternatives based on the best forecast performance.

The necessary revision of background error statistics turned out to be difficult indicating the sensitivity of the analysis-forecasting suite to small but systematic changes. The interdependence of 4DVAR, EDA, HRES and ENS will further increase in the future. The preparation of model upgrades will therefore become more cumbersome since performance must be evaluated across all components, between single model integrations and ensembles, and from short-to-monthly forecast ranges. The technical maintenance of the system increases as well and computational cost grows non-linearly with time. Coupling with ocean/sea-ice and including atmosperic composition will add further complexity.

Acknowledgements

Each model cycle contains a vast number of science based and technical modifications. The generation of new cycles is complex so that only about 2 new cycles can be implemented into operations per year. Many people across Research and Operations Departments at ECMWF contribute to this effort whose work is greatly acknowledged.

References

[1] A. I. Barrett. *Why can't models simulate mixed-phase clouds properly?* PhD thesis, Reading University, UK, 2012.

[2] A. Beljaars. Numerical schemes for parametrizations. Seminar Proceedings, 9-13 September 1991, Numerical methods in atmospheric models 130, ECMWF, 1992.

[3] M. Bonavita, L. Isaksen, and E. Hólm. On the use of EDA background error variances in the ECMWF 4DVAR. *Q. J. Roy. Meteor. Soc.*, 138:1540–1559, 2012.

[4] R. Buizza. Sensitivity of optimal unstable structures. *Q. J. Roy. Meteorol. Soc.*, 120:429–451, 1994.

[5] H. Grenier and C.S. Bretherton. A moist PBL parameterization for large-scale models and its application to subtropical cloud-topped marine boundary layers. *Mon. Wea. Rev.*, 129:357–377, 2001.

[6] L. Isaksen, M. Bonavita, R. Buizza, M. Fisher, J. Haseler, M. Leutbecher, and L. Raynaud. Ensemble of data assimilations at ecmwf. Technical Memorandum 636, ECMWF, Reading, UK, Available from: http://www.ecmwf.int/publications/, 2010.

[7] C. Jakob, E. Andersson, A. Beljaars, R. Buizza, M. Fisher, E. Gérard, A. Ghelli, P. Janssen, G. Kelly, A.P. McNally, M. Miller, A. Simmons, J. Teixeira, and P. Viterbo. The ifs cycle cy21r4 made operational in october 1999. Newsletter 87, ECMWF, 2000.

[8] M. Janisková and P. Lopez. Linearized physics for data assimilation at ecmwf. Technical Memorandum 666, ECMWF, Reading, UK, Available from: http://www.ecmwf.int/publications/, 2012.

[9] et al. Klein, S. A. Intercomparison of model simulations of mixed-phase clouds observed during the arm mixed-phase arctic cloud experimentt: single-layer cloud. *Q. J. Roy. Meteor. Soc.*, 135:979–1002, 2009.

[10] R.S. Lindzen and M. Fox-Rabinovitz. Consistent vertical and horizontal resolution. *Mon. Wea. Rev.*, 117:2575–2583, 1989.

[11] A. P. Lock. The numerical representation of entrainment in parameterizations of boundary layer turbulent mixing. *Mon. Wea. Rev.*, 129:1148–1163, 2001.

[12] F. Lott and M.J. Miller. A new subgrid-scale orographic drag parametrization: Its formulation and testing. *Q. J. Roy. Meteor. Soc.*, 123:101–127, 1997.

[13] L. Magnusson. Update of model climate diagnostics. RD Internal Memorandum 1250, ECMWF, 2012.

[14] J.-F. Mahfouf. Influence of physical processes on the tangent-linear approximation. *Tellus*, 51:147–166, 1999.

[15] D. F. Parrish and J. C. Derber. The National Meteorological Center's spectral statistical interpolation analysis system. *Mon. Wea. Rev.*, 120:1747–1763, 1992.

[16] I. Sandu, A. Beljaars, P. Bechtold, T. Mauritsen, and G. Balsamo. Why is it so difficult to represent stably stratified conditions in nwp models? *J. Adv. Mod. Earth Syst.*, page in press, 2013.

[17] I. Sandu and B. Stevens. On the factors modulating the stratocumulus to cumulus transitions. *J. Atmos. Sci.*, 68:1865–1881, doi:10.1175/2011JAS3614, 2011.

[18] B. Stevens, A. Beljaars, S. Bordoni, C. Holloway, M. Köhler, S. Krueger, V. Savic-Jovcic, and Y. Zhang. On the structure of the lower troposphere in the summertime stratocumulus regime of the northeast pacific. *Mon. Wea. Rev.*, 135:985–1005, 2007.

[19] J. Teixeira. The impact of increased boundary layer vertical resolution on the ECMWF forecast system. Technical Memorandum 268, ECMWF, Reading, UK, Available from: http://www.ecmwf.int/publications/, 1999.

[20] Y. Trémolet. Model-error estimation in 4D-Var. *Q. J. Roy. Meteor. Soc.*, 133:1267–1280, 2007.

Appendix

Table 2 lists the experiment identification of the main experiments used for impact evaluation of cycle 38r2. Figures 46-52 show scorecard summaries od experiment performance.

Table 2: Experiments

	Experiment (38r2)	Control (38r1)	Period	Comment
EDA:				
	fsqg	0058	20120101-20120331	Uses fsqs 4DVAR experiment for filtering/calibration
	fsqh	0001	20120601-20121231	Uses fsqu 4DVAR experiment for filtering/calibration
HRES	initial:			
	folg	N/A	20110603-20110901	L91 forecasts, initialized with own L91 analyses
	fr0n	N/A	20110603-20110630	L91 forecasts, initialized with fq6k L137 analyses
	fr2s	N/A	20110601-20110713	L137 forecasts, initialized with own L137 analyses
	fqy1	N/A	20110603-20110630	L137 forecasts, initialized with folg L91 analyses
HRES	final:			
	fsr8	0058	20120101-20120331	Uses fsqg EDA, initialized at 0/12 UTC, 12 hours between forecasts
	fsra	0001	20120601-20121231	Uses fsqh EDA, initialized at 0/12 UTC, 12 hours between forecasts
ENS:				
	ftf0	ftan	20120114-20120314	Initialized at 0 UTC, 48 hours between forecasts
	ftg7	0001	20120620-20120813	Initialized at 0 UTC, 48 hours between forecasts
	0062	0001	20130110-20130317	Initialized at 0 UTC, 24 hours between forecasts
	ftm3	focq	01021989-15122008	45-day integrations, every 3 months

Scorecard

dates=[2011060200,2011060300,2011060400,...,2011062900,2011063000]
steps=[12, 24, 36, 48, 60, 72, 84, 96, 108, 120, 132, 144, 156, 168, 180, 192, 204, 216, 228, 240]
reftypes=['ob']
refstreams=['oper']
streams=['oper']
expvers=(cntrl:['folg'], exper:['fsqi'])
refexpvers=['0001']
refclasss=['od']
vstreams=['NewVerPB_38R2']
types=['fc']
classs=['rd']
normalisation_method=mean
confidence=[95.0]

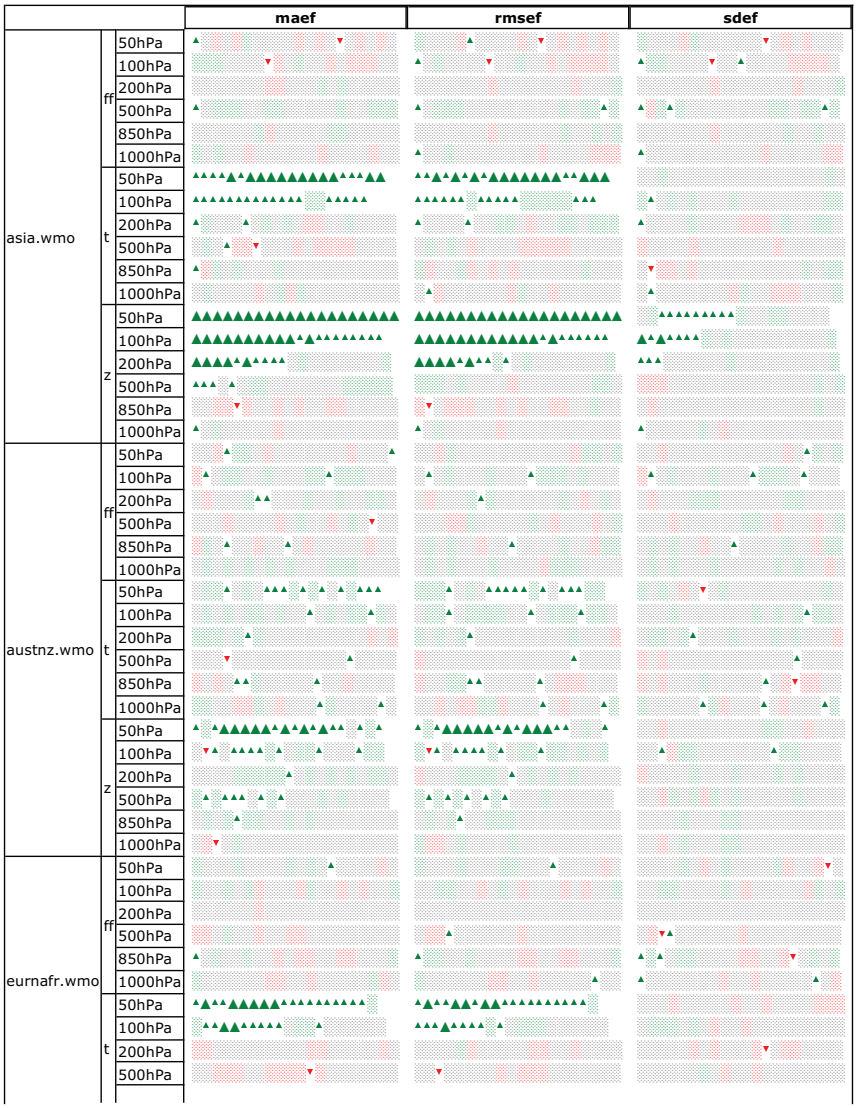


Figure 46: Scorecard from T511 experiment evaluation against observations showing 10-day forecast skill improvement (green) / deterioration (red) for 38r2 against 38r1, both with 91 levels. Triangles indicated areas where score differences are statistically signifcant to the 95% level. maef: mean error, rmsef: root-mean-square error, sdef: error standard deviation.

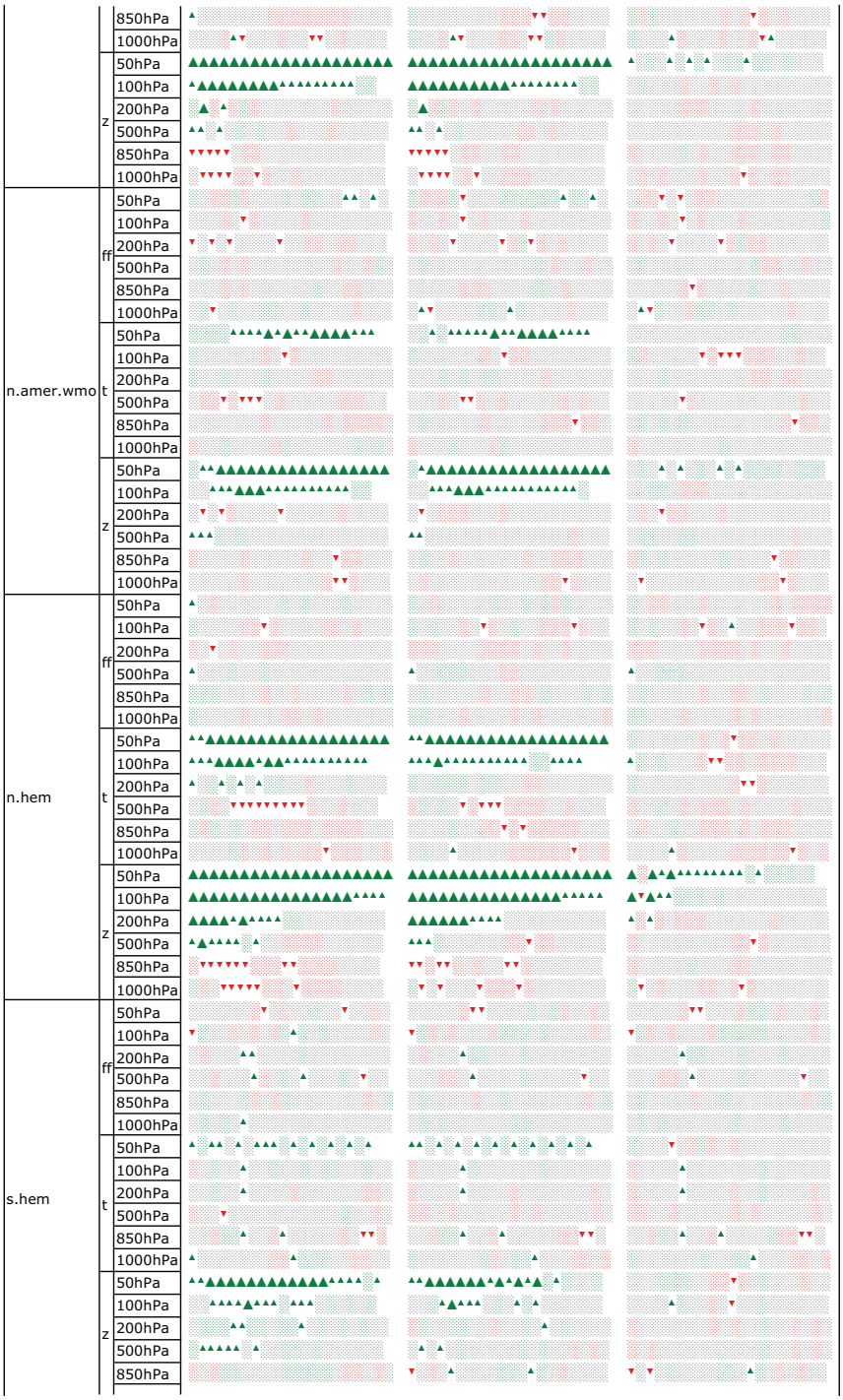
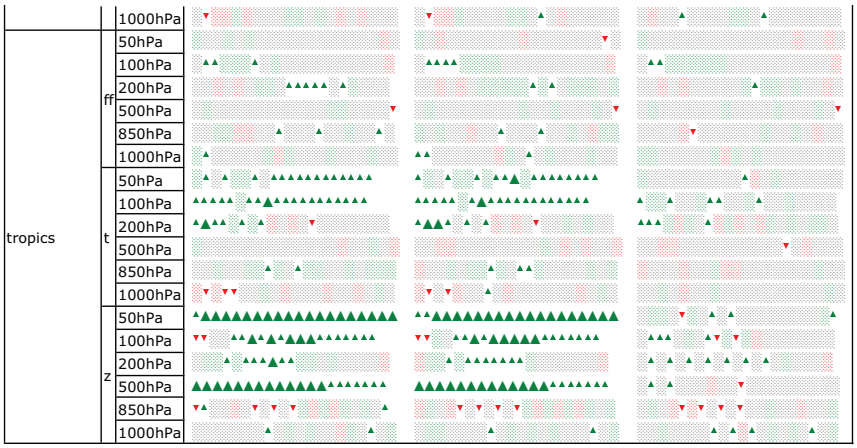


Figure 47: Fig. 46 continued.



Score card provides a quick visual overview over the performance of the experiment scores compared to control. It is a simplified summary of verify error plots of various domains, scores, parameters etc.
Each error plot is converted into a sequence of symbols (e.g. ▼▾▾ ▴▲▲) where each symbol indicates for given time step whether or not the experiment is significantly better or worse than the control.

Symbol legend: for a given forecast step... (d: score difference, s: confidence interval width)
▲ experiment **better** than control **statistically highly significant** (the confidence bar above zero by more than its height) (d/s>3)
▴ experiment **better** than control **statistically significant** (d/s≥1)
experiment better than control, yet not statistically significant (d/s≥0.5)
not really any difference between control and experiment
experiment worse than control, yet not statistically significant (d/s≤-0.5)
▾ experiment **worse** than control **statistically significant** (d/s≤-1)
▼ experiment **worse** than control **statistically highly significant** (the confidence bar below zero by more than its height) (d/s<-3)

Figure 48: Fig. 47 continued.

Scorecard

dates=[2012010100,2012010112,2012010200,...,2012122100,2012122112]
steps=[12, 24, 36, 48, 60, 72, 84, 96, 108, 120, 132, 144, 156, 168, 180, 192, 204, 216, 228, 240]
reftypes=an
refstreams=oper
streams=oper
expvers=(cntrl:['0001', '0058'], exper:['fsr8', 'fsra'])
refexpvers=(cntrl:0001, exper:fsra)
refclasss=(cntrl:od, exper:rd)
vstreams=(cntrl:NewVerPB_38R2, exper:['NewVerPB_38R2', 'NewVerPB_38R3'])
types=fc
classs=(cntrl:od, exper:rd)
normalisation_method=mean
confidence=[95.0]

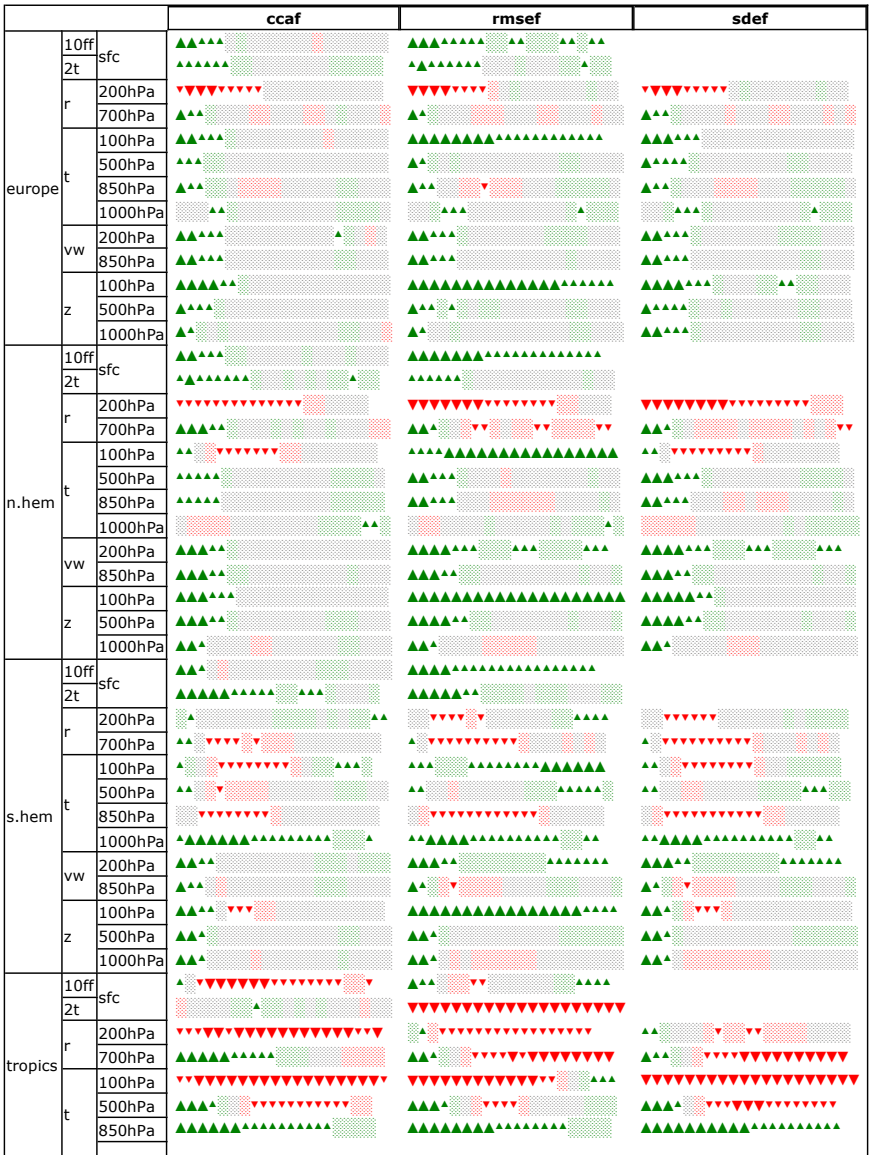
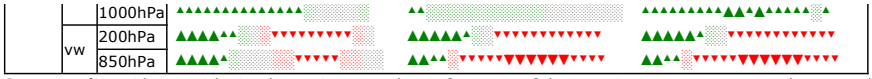


Figure 49: Scorecard from RD e-suite evaluation against own analysis showing 10-day forecast skill imporvement (green) / deterioration (red) for 38r2 against 38r1 control suites. Triangles indicated areas where score differences are statistically signifcant to the 95% level. ccaf: anomaly correlation, rmsef: root-mean-square error, sdef: error standard deviation.



Score card provides a quick visual overview over the performance of the experiment scores compared to control. It is a simplified summary of verify error plots of various domains, scores, parameters etc.
Each error plot is converted into a sequence of symbols (e.g. ▼▼▾▾▴▲▲) where each symbol indicates for given time step whether or not the experiment is significantly better or worse than the control.

Symbol legend: for a given forecast step... (d: score difference, s: confidence interval width)
▲ experiment **better** than control **statistically highly significant** (the confidence bar above zero by more than its height) (d/s>3)
▴ experiment **better** than control **statistically significant** (d/s≥1)
▒ experiment better than control, yet not statistically significant (d/s≥0.5)
▒ not really any difference between control and experiment
▒ experiment worse than control, yet not statistically significant (d/s≤-0.5)
▾ experiment **worse** than control **statistically significant** (d/s≤-1)
▼ experiment **worse** than control **statistically highly significant** (the confidence bar below zero by more than its height) (d/s<-3)

Figure 50: Fig. 49 continued.

Scorecard

dates=[2012010100,2012010112,2012010200,...,2012123100,2012123112]
steps=[12, 24, 36, 48, 60, 72, 84, 96, 108, 120, 132, 144, 156, 168, 180, 192, 204, 216, 228, 240]
reftypes=ob
refstreams=oper
streams=oper
expvers=(cntrl:['0001', '0058'], exper:['fsr8', 'fsra'])
refexpvers=0001
refclasss=od
vstreams=(cntrl:NewVerPB_38R2, exper:['NewVerPB_38R2', 'NewVerPB_38R3'])
types=fc
classs=(cntrl:od, exper:rd)
normalisation_method=mean
confidence=[95.0]

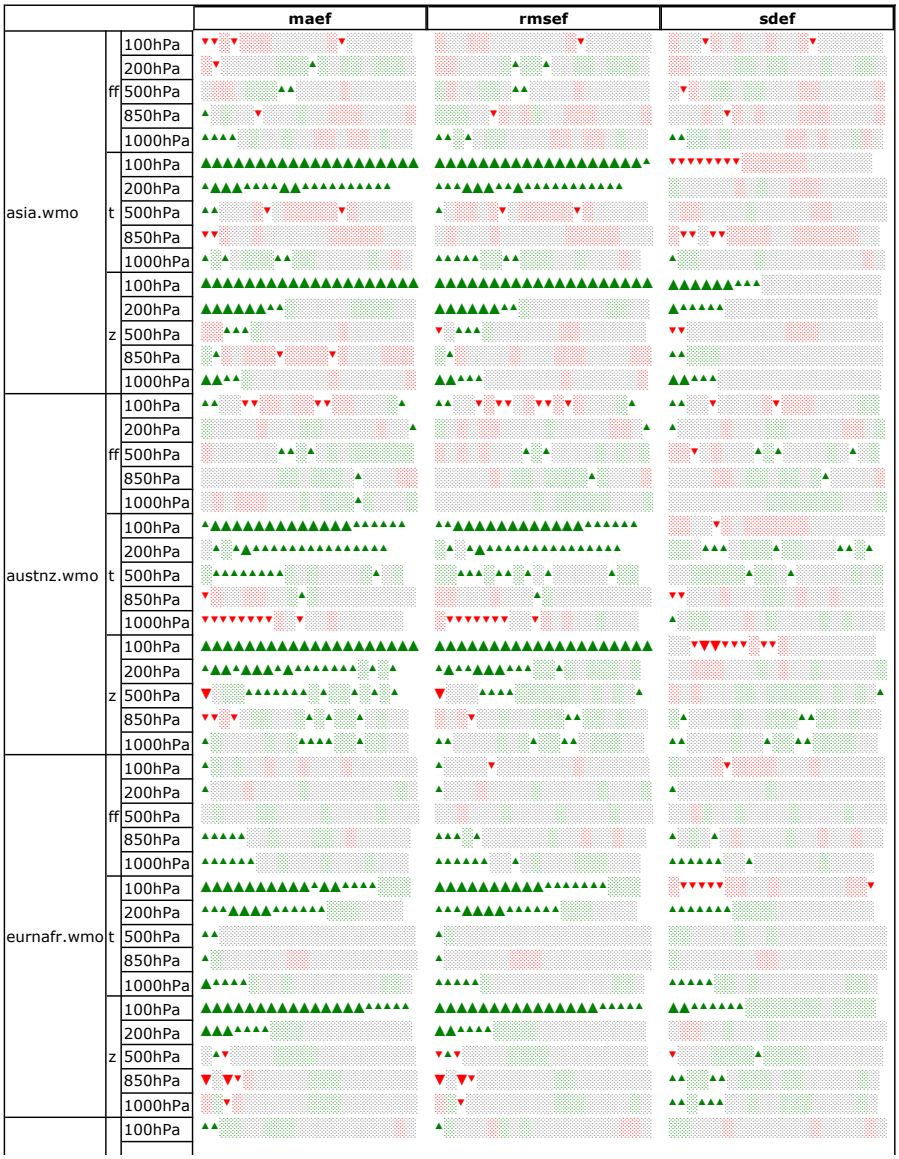
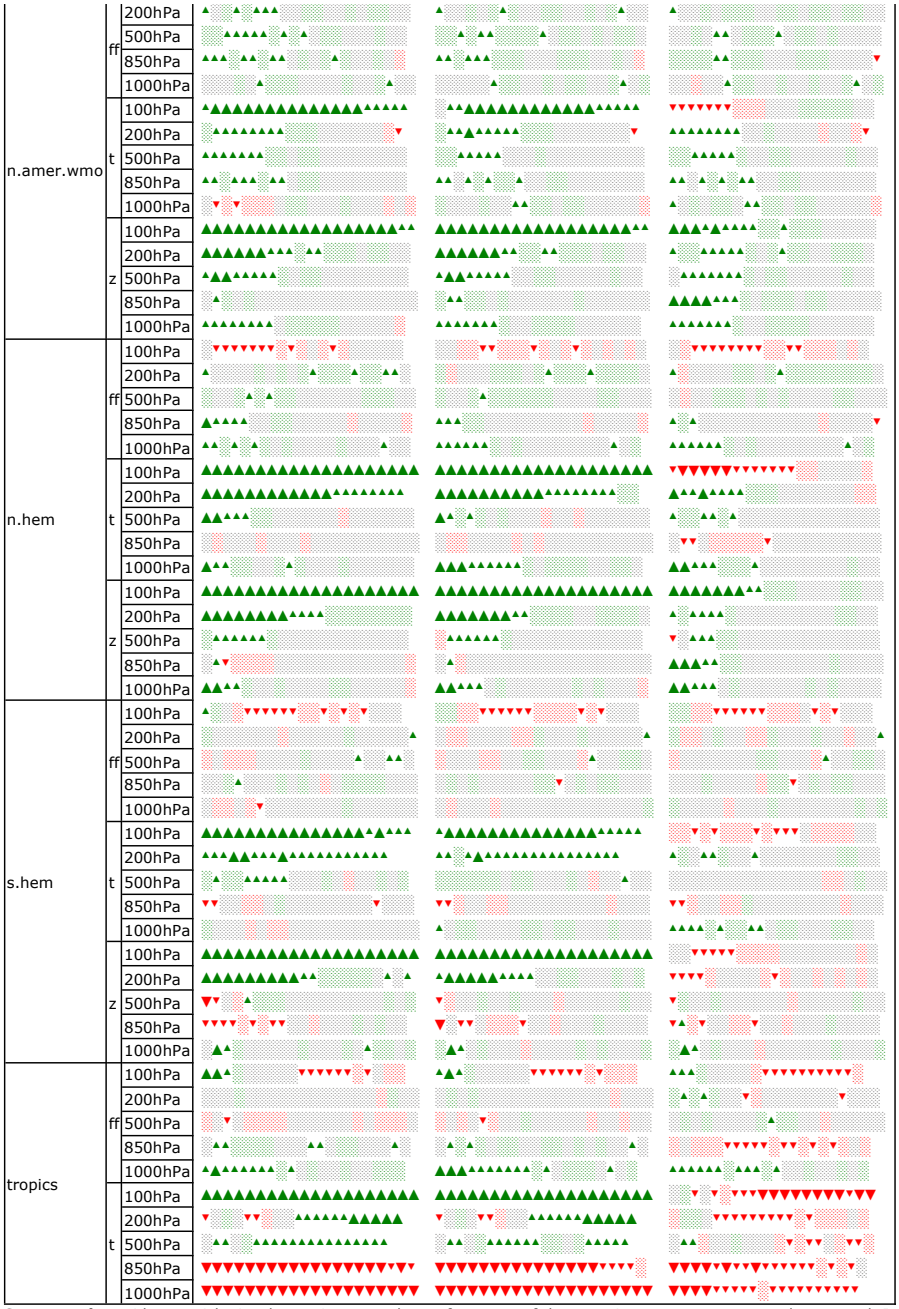


Figure 51: Scorecard from RD e-suite evaluation against observations showing 10-day forecast skill improvement (green) / deterioration (red) for 38r2 against 38r1 control suites. Triangles indicated areas where score differences are statistically significant to the 95% level. msef: mean error, rmsef: root-mean-square error, sdef: error standard deviation.



Score card provides a quick visual overview over the performance of the experiment scores compared to control. It is a simplified summary of verify error plots of various domains, scores, parameters etc.
Each error plot is converted into a sequence of symbols (e.g. ▼▼▼ ▲▲▲) where each symbol indicates for given time step whether or not the experiment is significantly better or worse than the control.

Symbol legend: for a given forecast step... (d: score difference, s: confidence interval width)
▲ experiment **better** than control **statistically highly significant** (the confidence bar above zero by more than its height) (d/s>3)

Figure 52: Fig. 51 continued.