Seventh Framework Programme
Theme 6 [SPACE]

UERRA
Uncertainties in Ensembles of Regional ReAnalyses

Project: 607193 UERRA

Full project title:
Uncertainties in **E**nsembles of **R**egional **Re-A**nalyses

Final Report

WP no:	2
WP leader:	MO
Lead beneficiary for deliverable :	MO
Name of author/contributors:	Peter Jermey, Amy Doherty and Richard Renshaw
Nature:	Report
Dissemination level:	PU
Deliverable month:	48
Submission date: 20th December 2017	Version nr: 1

Uncertainties in Regional Reanalyses
Final Report (Met Office)

Peter Jermey, Amy Doherty and Richard Renshaw

December 20, 2017

1 Introduction

The principal contributions of the Met Office to the Uncertainties in Regional Reanalysis (UERRA) project, [Unden et al., 2014], are a deterministic atmospheric reanalysis at 12km and a 20 member ensemble atmospheric reanalysis at 36km over Europe for the satellite era (1979-2016). Production started in January 2017 and is expected to complete in January 2018. The development of the project is documented in deliverable reports, which are publicly available through the UERRA website, *http://uerra.eu*. In D2.1, [Jermey et al., 2015], the reanalysis systems are documented, together with initial test results and a demonstration of ensemble products. D2.2, [Mahmood et al., 2016], describes use of observations and other data used as boundary conditions for the reanalyses. Initial diagnostics from the ensemble system are presented in D2.3, [Jermey et al., 2016], which demonstrates the validity and quality of the ensemble production data, as well as reporting on trials of the satellite bias correction method, land surface analysis scheme and ensemble size. The deterministic reanalysis uses a hybrid approach to data assimilation, exploiting the ensemble information to improve the reanalysis, as documented in D2.4, [Jermey et al., 2017b]. D2.4 also contains tuning and initial trial results for the deterministic reanalysis system. An assessment of the quality of the ensemble reanalysis is given in D2.14, [Jermey et al., 2017a], which includes comparison with the 20th century ensemble reanalysis (CERA-20C), [Laloyaux et al., 2016] from the European Centre for Medium Range Weather Forecasts (ECMWF) and the University of Bonn/Deutscher Wetterdienst ensemble reanalysis for UERRA, [Bach et al., 2016].

This final report summarises and updates the documentation laid out in the deliverables and presents evaluation of both ensembles for the entire period, 1979-2016. Documentation of the reanalysis systems is summarised in section 2. Evaluation results for the deterministic reanalysis and the ensemble reanalysis are presented in sections 3 and 4, respectively. The impact of satellite data assimilation on the uncertainty in the ensemble reanalysis is examined in section 5. Finally, archiving information is given in section 6 and a summary is given in section 7.

Period	Deterministic		Ensemble
Resolution	0.11° ≈ 12km		0.33° ≈ 37km
Assimilation Resolution	0.22° ≈ 24km		0.66° ≈ 75km
Grid size	432 x 432		144 x 144
NW Corner	60.3°N, −45.4°E,	near Agdleruussakasit, Greenland	Same
SW Corner	20.2°N, −9.5°E,	near Adrar, Mauritania	Same
SE Corner	23.2°N, 36.5°E,	near Shalateen, Egypt	Same
NE Corner	66.6°N, 66.7°E,	near Salekhard, Russia	Same
Pole	39.25°N, 198.0°E		Same
Number of Levels	63 levels to 40km		Same

Table 1: Rotated pole grid specification.

2 Documentation

2.1 Scope

The reanalyses are driven by advanced data assimilation techniques which produce an estimate of the atmospheric state, by drawing a background atmospheric state towards a set of observations. These reanalyses are produced every six hours, with an atmospheric model forecasting to the next reanalysis time. This (re)forecast is then the background atmospheric state for the next reanalysis. The reanalyses cover the EUR-11 domain of the Coordinated Regional climate Downscaling Experiment (CORDEX, [Jacob et al., 2014]), as shown in figure 1. The grid specification for both reanalyses is given in table 1.

Both reanalyses cover the period 1979-2016 and are free cycling, but are dependent on ERA-Interim for lateral boundary conditions (LBCs). Each production stream also requires an initial background to start the cycling process. This data also comes from ERA-Interim. Each stream has a month (ensemble) or a week (deterministic) of spin-up time from this initial background before production data is processed and archived. As an example, the ensemble stream of 1997-1999 is spun-up from ERA-Interim data on 1st December 1996 and begins producing and archiving data on the 1st January 1997. Since ERA-Interim data is not available prior to 1979, the exception to this pattern is the first month of 1979, which has not been spun-up and therefore will be more heavily influenced by the parent model than subsequent data. This restriction also means that the first six hours of 1979 are not produced (00UTC on 19790101 to 05UTC on 19790101).

2.2 Observations

The observations assimilated by the reanalyses are mostly supplied by European Centre for Medium Range Weather Forecasts (ECMWF), [Dee et al., 2011]. Prior to 2003 these observations come from ECMWF's reanalysis archives and 2003-2016 from operational archives. These

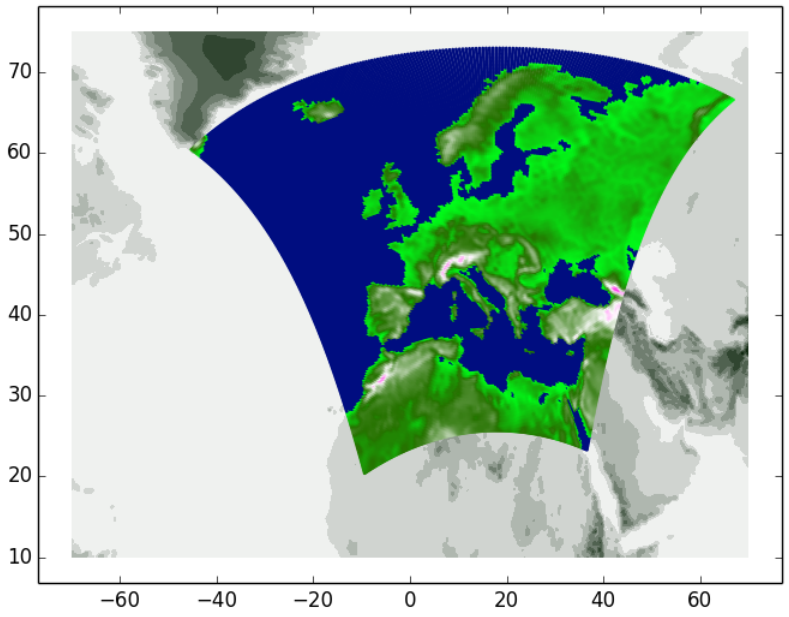

Figure 1: Reanalysis domain.

observations include SYNOP stations, ships, buoys, sondes, aircraft and satellite measurements using advanced infared sounders (AIRS), (Advanced) Television Infrared Observation Satellite Operational Vertical Sounder ((A)TOVS), scatterometer winds and Infrared Atmospheric Sounding Interferometer (IASI). These are complemented with atmospheric motion vectors (AMV) from the Met Office and reprocessed global positioning system zenith total delays (GroundGPS), [Pacione et al., 2017]. Due to a technical error only GroundGPS data from 2014-2015 is assimilated. Further satellite data using GPS radio occulation (GPSRO), Spinning Enhanced Visible and InfraRed Imager clear sky (SEVIRIclear) and Special Sensor Microwave Imager/Sounder (SSMIS) was not attempted due to time and resource limitations. A guide to the number of assimilated observations is given in figure 2. This figure shows that there is a substantial increase in surface, aircraft and satellite observation volume between 1979 and 2016, but sonde volume remains relatively steady.

The new Monitoring and Updating Station Lists (MUSLi) system is used to reject and correct observations based on monthly departure statistics. Corrections are made for surface observations of pressure and for temperature observations from aircraft and upper air data, [Davie, 2017].

Variational bias correction is applied to the satellite radiances, [Lorenc, 2012, Dee and Uppala, 2009]. This allows for bias corrections to vary with time so as to fit drifts in instrument bias. The method has been previously well tested for global models and is used here for the first time in

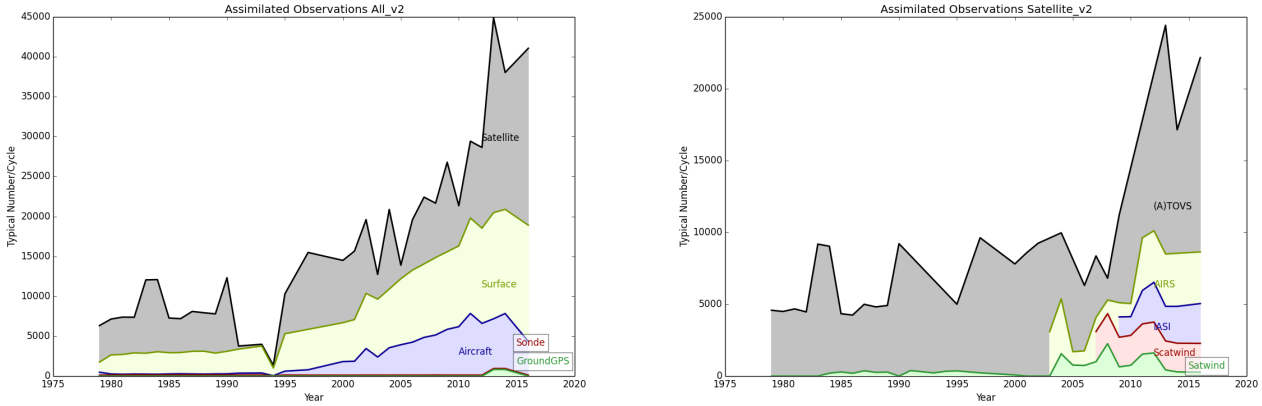

Figure 2: Guide to the number of each type of observation assimilated by the reanalyses. LHS all observations, RHS satellite data.

a regional model. High peaking channels are rejected, see [Jermey et al., 2015].

2.3 Boundary Conditions

For LBCs, it was originally intended to use ERA5 (31km horizontal resolution, 137 levels to 80km), [Dee, 2014], since this includes a 10 member ensemble, which would have supplied the ensemble reanalysis with perturbed boundary conditions. Due to time constraints, it was not possible to use this dataset and instead, deterministic LBCs were used from ERA-Interim (60km horizontal resolution, 60 levels to 80km), [Dee et al., 2011].

The upper boundary of the model is a rigid lid at approximately 40km, which is lower than is typical for a global model, [Brown et al., 2012].

For most of the period, the lower sea boundary is given by version 2 of the Hadley Centre Sea Ice and Sea Surface Temperature (HadISST2), [Titchner and Rayner, 2014]. This contains a ten member analysis of sea surface temperatures at 0.25° consistent with a single deterministic analysis of sea ice fraction at 1.0°. This dataset does not cover 2011-2016, so from September 2011, the Met Office Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA), [Donlon et al., 2012], is used, reduced to the resolution of HadISST2. For this later period, an ensemble dataset is produced by adding HadISST2 from a random year to the reduced resolution OSTIA fields. During both periods, the ten member ensemble of sea boundary data is shared so that each sea boundary member is used by two reanalysis ensemble members. The lower sea boundary is detailed in table 2.

The land boundary is provided by the Met Office Land Surface Data Assimilation System

Period	Deterministic	Ensemble perturbations
1979-2010	HadISST2 mean	HadISST2 perturbations
2010-2016	OSTIA	HadISST2 perturbations using random year (1979-2010)

Table 2: Table displaying lower sea boundary for deterministic and ensemble reanalyses.

(SURF), [Candy, 2014], used in a regional context for the first time. As with the atmospheric analysis, each member of the ensemble performs its own land surface analysis from a different realisation of in-situ observations.

2.4 Ensemble System

In order for the ensemble of reanalyses to represent uncertainty in the reanalysis system, the sources of uncertainty are perturbed so that each member receives a different realisation. In addition to different realisations of the boundaries, different realisations of the model and observations are also employed. The model is perturbed by adding random analysis increments as tendencies across the forecast. A random analysis increment may be thought of as a proxy for model error. Each observation is also perturbed by adding a random amount bounded by the estimated observation error. This follows a system for generating global ensembles used at the Met Office for research into model error, [Piccolo and Cullen, 2016]. Each ensemble member is isolated as there is no re-centring of the ensemble at analysis time. An unperturbed control member is run separately from the ensemble in order to monitor and update observation blacklists and calculate satellite bias correction at the ensemble resolution.

The ensemble reanalysis was originally intended to use four-dimensional variational assimilation (4DVAR), [Rawlins et al., 2007]. However this proved to be prohibitively expensive and so three-dimensional variational assimilation (3DVAR), [Lorenc et al., 2000] is used instead. 3DVAR suffers from spin-up issues for precipitation so a consequence of this change is that the ensemble precipitation values are far too large. The spread of the precipitation is, however, still useful for estimating reanalysis uncertainty, [Jermey et al., 2017a].

Variational assimilation (3DVAR, 4DVAR and hybrid 4DVAR) draws the model close to the observation data, which in the case of the ensemble is perturbed, by minimising a cost function which penalises distance from observations, distance from the background and high frequency behaviour. The weights to the observation and background terms are inverses of the estimates of the error covariances. It produces an increment to the background which estimates the optimum state of the atmosphere, given the background and observations across a six hour assimilation window (T-3 to T+3), taking into account the three-dimensional position of each observation. As is common practice, the assimilation is carried out at half the horizontal resolution of the model.

2.5 Deterministic System

The deterministic reanalysis uses hybrid 4DVAR, [Clayton et al., 2013], to assimilate observations. As with 3DVAR, this is carried out by minimising a cost function, which penalises distance from observations, distance from the background and high frequency behaviour. (Hybrid) 4DVAR also estimates the optimum state of the atmosphere, given the background and observations across a six hour assimilation window (T-3 to T+3), taking into account the three-dimensional position of each observation, but also taking into account the validity of each observation within the window. The use of the fourth dimension leads to analysis increments which are more in balance with the model than 3DVAR and therefore it does not suffer from precipitation production which is initially too large.

Traditionally the background error covariance is estimated by a smooth parameterised approximation to climatology tuned by forecast differences over a long period. This is sufficiently accurate to consistently produce an analysis which is closer to the true atmospheric state than the background. However, this estimation does not vary from cycle to cycle and regions of activity in the background are given the same weight as regions of stability.

Hybrid 4DVAR uses a combination of the climatological background error covariance (used for the ensemble) and a error covariance derived from the ensemble. This combined covariance creates a low-noise estimation that is dependent on synoptic pattern. This assimilation method has been used for background error covariance estimation in operational global forecasting at the Met Office since 2011, [Clayton et al., 2013]. It is used here for the first time in a regional context. A diagram of this is shown in figure 3 and the hybrid settings are detailed in table 3, which were determined by a combination of experimental tuning, [Jermey et al., 2017b], and in-house experience. To avoid spuriously large covariances near the top of the atmosphere, climatology only is used for this region. A middle region features a linear ramp from hybrid covariances to climatological covariance. Use of the ensemble data ensures that the background error covariance is synoptically dependent in the deterministic reanalysis.

2.6 Model

The same model is used by both the ensemble and the deterministic reanalyses. This is the Unified Model (UM), [Davies et al., 2005], using the dynamical core currently used for operational forecasting at the Met Office (Even Newer Dynamics for General atmospheric modelling of the environment, ENDGame, [Wood et al., 2014]). The reanalyses both use six hour cycling, following the Met Office operational global forecasting system. A (re)forecast is carried out from a combination of the analysis increment, provided by the data assimilation, and the previous forecast (background) from T-3 to at least T+9. The reanalysis fields are output at T+0 (the centre of the six hour assimilation window) and the background for the next cycle between T+3 and T+9. The model is computed over 63 levels using Charney-Phillips staggering between 10m above orography to the model top - a fixed radius from the centre of the Earth.

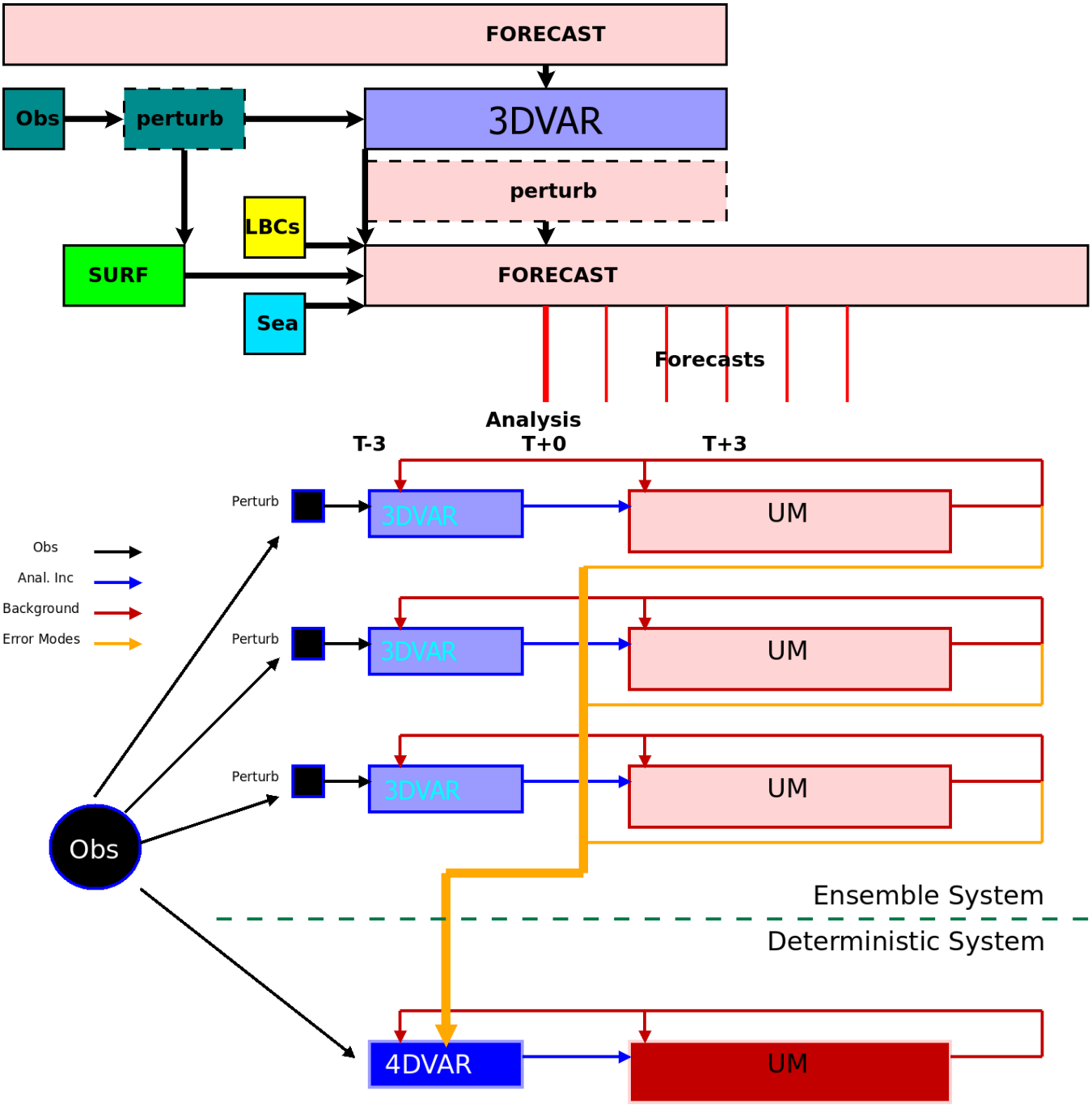

Figure 3: System design. Top shows a single member of the ensemble to demonstrate system cycling. Bottom shows interaction between ensemble and deterministic system. Only three members of the ensemble system are shown.

Item	Value
Upper Zone: Climatology covariance weight	100%
Upper Zone: Ensemble covariance weight	0%
Lower Zone: Climatology covariance weight	100%
Lower Zone: Ensemble covariance weight	30%
Start of middle zone	16km
Start of upper zone	21km
Vertical localisation pressure change factor	1.5
Remove scales above	282km
Ensemble size	20 members
Horizontal localisation scale	200km

Table 3: Hybrid settings. The localisation uses the Gaspari-Cohn function, [Gaspari and Cohn, 1999], to reduce noise in the covariances.

2.7 Suite Design

For each production stream, three suites are run, control, ensemble and deterministic. The control (a deterministic system at ensemble resolution, not archived) prepares necessary files for the ensemble including satellite bias correction. The three types of suite use the same design with different optional tasks. The suites are designed and run using the Met Office Rose system, [Various, 2017], based on the New Zealand National Institute of Water and Atmospheric Research's cylc system, [Oliver, 2017]. This schedules tasks based on predefined dependencies allowing for automatic production. A summary of the tasks run by the suites is given in tables 4 and 5.

3 Deterministic Results

3.1 Assimilation Statistics

Figure 4 shows some of the assimilation statistics for the deterministic reanalysis for the month of March from selected years. The left hand column shows the number of assimilated sondes and the central column shows the number of assimilated temperature observations from surface stations. This shows that the number of assimilated surface observations increases greatly over the period of the reanalysis from of the order of 1000 in 1979 to approximately 14000 in 2016, whereas the number of assimilated sondes remains reasonably constant, of the order of 100 at 00Z and 12Z, with fewer at 06Z and 18Z. The plots show that these counts are mostly stable over the cycles shown except for a diurnal cycle. Each plot shows a period where there is a dramatic drop in the number of assimilated surface observations, which is due to a decrease in available observations in the archive. Some of these are reflected in a decrease in the number of

Task	/Ob	/Mem	C	E	D	Notes
obstype_selector		x	x	x		Select observation sources
ra_ops_process_	x	x	x	x	x	Process observations for assimilation
ra_ops_process_anal	x	x	x	x	x	Add assimilation statistics to ODBs
ra_create_obstore_	x		x		x	Convert from ODB1 to ObStore
ra_ops_odb_to_odb2_	x	x	x	x	x	Convert from ODB1 to ODB2
ra_ops_bufr_odb_	x		x	x	x	Convert from BUFR to ODB1
update_biases_groundgps		x	x	x		Update biases for ground GPS
perturb_varobs_	x	x		x		Perturb observations
check_sst		x	x	x		Check that SST fields are available
prepare_surf		x	x	x		Prepare surface boundary fields
ra_surf_anal_sst		x	x	x	x	Prepare sea boundary
varbc_selector				x		Set satellite biases
ra_surf_um2jules		x	x	x	x	Convert from atmospheric to land model
ra_surf_jules		x	x	x	x	Surface model
ra_surf_ekf		x	x	x	x	Surface analysis
ra_var_anal		x	x	x	x	Assimilation
ra_var_anal_screen		x	x	x	x	Assimilation for surface analysis
ra_var_bc			x		x	Bias Correction calculation for satellites
perturbation_selector_		x		x		Select perturbation for model
ra_um_recon_daily		x	x	x	x	Reconfigure daily fields
ra_um_fcst		x	x		x	Atmospheric model
ra_um_fcst_long					x	Longer version of atmospheric model
ra_um_recon_ls		x	x	x	x	Reconfigure fields for assimilation
ra_um_recon_ls_screen		x	x	x	x	Reconfigure fields for surface analysis

Table 4: Principal tasks in production systems. Some tasks are run multiple times for each cycle, either per observation type (/Ob) or per member (/Mem) or both. Tasks may be used for the control (C), the ensemble (E) and/or the deterministic (D) reanalysis.

Task	/Ob	/Mem	C	E	D	Notes
gen_ret_month			x	x	x	Retrieve monthly files from archives
gen_ret_cycle					x	Retrieve cycle files from archives
cycle_check			x	x	x	Check satellite channels
cycle_install			x	x	x	Change satellite channels
housekeep			x	x	x	Clean up general files
housekeep_odb2			x	x	x	Clean up ODB2 files
housekeep_verodb			x	x	x	Clean up ODB1 files
gen_ver_hk_fdb			x	x	x	Clean up verification files
gen_ver_hk_ard			x	x	x	Clean up verification files
sat_stats_tidy					x	Tidy satellite statistics
ra_archive		x	x	x	x	Temporary archive of files
ra_archive_monthly			x	x	x	Temporary archive monthly files
ra_archive_hybridprep				x		Archive error modes for hybrid
pack_logs			x	x	x	Tidy log files
check_local_horse			x	x	x	Tidy processing software and files
ff2grib			x	x	x	Convert from fieldsfile to grib2
ppra_		x	x	x	x	Process output files
mars_control			x		x	Archive to mars (deterministic)
mars_ensemble				x		Archive to mars (ensemble)
recon_em_src				x		Reconfigure for error modes
create_error_modes				x		Create error modes for hybrid
suite_stats	x	x	x		x	Monitor assimilation statistics
musli_setup	x		x		x	Setup observation blacklist monitoring
musli_decision_	x		x		x	Create blacklist decision
musli_make	x		x		x	Make blacklist
musli_monitor	x		x		x	Monitor blacklist
ra_ver_model_fdb			x		x	Convert fields to verification format
ra_ver_obs_	x		x		x	Verify against observations
engl_ver_model_fdb		x		x		Convert fields to verification format
engl_ver_obs_	x			x		Verify against observations
engl_ver_anl				x		Verify against mean (spread)

Table 5: Archiving, housekeeping and monitoring tasks in production systems. Some tasks are run multiple times for each cycle, either per observation type (/Ob) or per member (/Mem) or both. Tasks may be used for the control (C), the ensemble (E) and/or the deterministic (D) reanalysis.

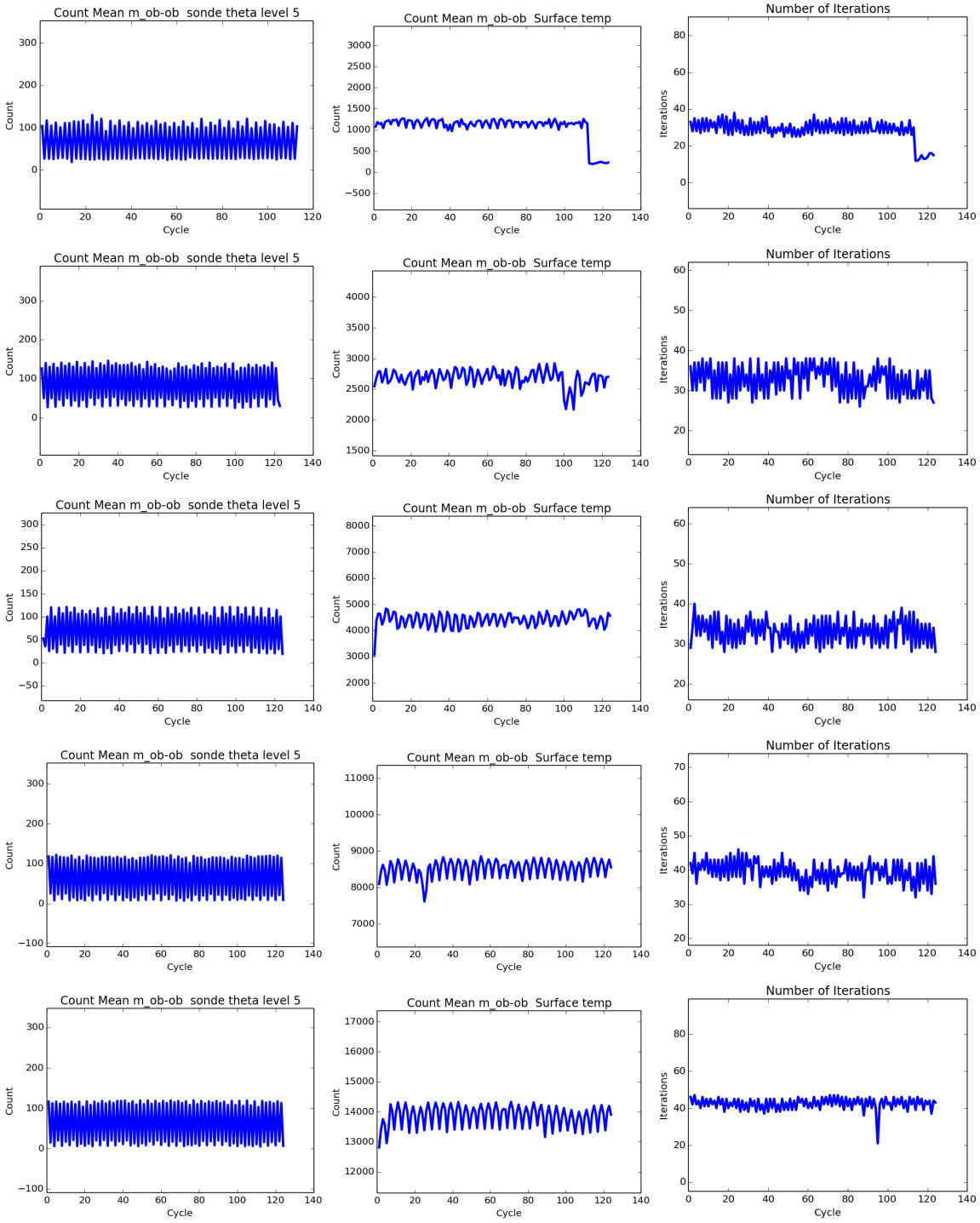

Figure 4: Assimilation statistics. LHS is the number of assimilated sondes. Central is the number of assimilated temperature observations from surface stations. RHS is the number of iterations of VAR to convergence. Top to bottom: March 1979, 1988, 1997, 2006 and 2016.

iterations required for convergence in the assimilation, shown in the right hand side plots. The plot for 2016 shows a single cycle with a much lower iteration count than the rest. On this cycle the convergence failed and the system automatically returned a increment of zero.

Figure 5 shows the RMS difference between pressure observations from surface stations and the background (O-B) and between these observations and the reanalysis (O-A). They also show O-A for sonde observations of potential temperature at model level 5. Again these statistics are for March from selected years of the deterministic reanalysis. This figure shows that the reanalyses are closer to the observations than the background and that the distances of the background and reanalyses from the observations generally decreases with time. This shows that the assimilation process correctly draws the model towards the observations and that increasing the volume of observations reduces the error in both the background and the reanalyses. The reduction in the number of assimilated observations towards the end of March 1979 manifests as an increase in both statistics. The assimilation bust in March 2016 is not noticeable, indicating that it did not cause a large reduction in the quality of the reanalysis for subsequent cycles. The two peaks seen in the surface pressure statistics for March 2006 are likely related to the synoptic situation and are not seen in other assimilation statistics apart from the cost function terms which penalise high frequency behaviour.

Futher assimilation statistics are presented in D2.4, [Jermey et al., 2017b].

3.2 Reanalysis Quality

It is difficult to assess the reanalysis quality directly due to the lack of availability of independent measures of truth. This is because most available high quality measurements of atmospheric variables are assimilated into the reanalyses. It is common instead to assess the six hour forecast from the reanalysis against observations (the background). Since errors grow with forecast time these assessment metrics are an upper bound on the true errors in the reanalysis.

Figure 6 shows the RMSE and mean error of 6h forecasts from the UERRA-MO deterministic reanalysis and ERA-Interim against 2m temperature observations for March 1979, 1988, 1997, 2006 and 2016. This demonstrates that the quality of the representation of 2m temperature in UERRA-MO is an improvement on ERA-Interim throughout the reanalysis period. This benefit increases with increasing time/observation volume, see table 6, which details the percentage change in RMSE from ERA-Interim to UERRA-MO. This suggests that higher resolution assimilation is of greatest benefit when a dense observation network is available. The quality of both reanalyses improves with time and both reanalysis have low mean error (bias) throughout, with UERRA-MO slightly warmer than ERA-Interim.

Figure 7 shows the RMSE and mean error of 6h forecasts from the UERRA-MO deterministic reanalysis and ERA-Interim against 10m wind vector observations for the same periods as figure 6. As with 2m temperature, both reanalyses improve with time/observation volume. Again,

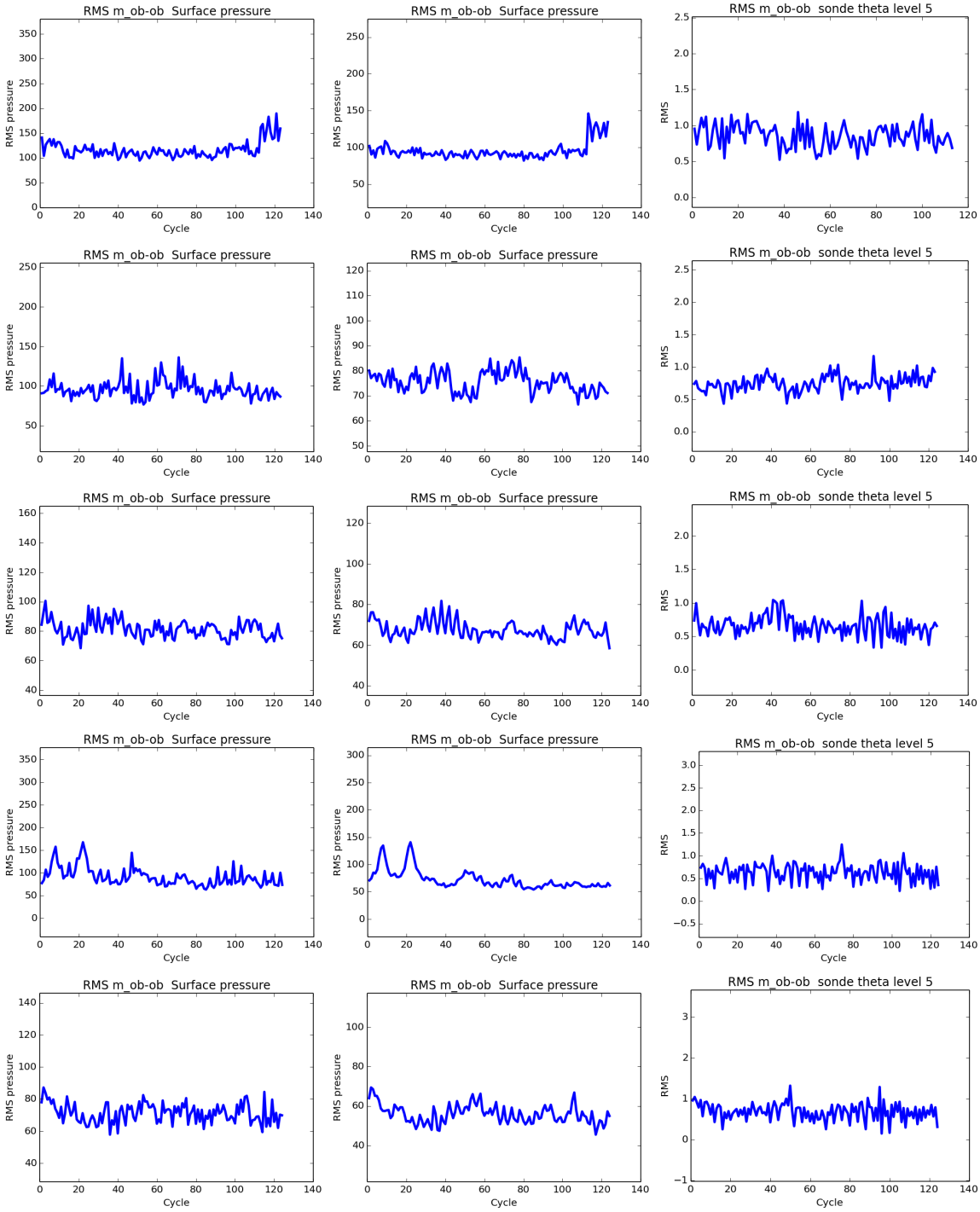

Figure 5: Assimilation statistics. LHS is the RMS difference between pressure observations from surface stations and the background (O-B). Central is the RMS difference of these observations with the reanalysis (O-A). RHS is the RMS difference of the reanalyses with sonde observations of potential temperature at model level 5 (approximately 980hPa). Top to bottom: March 1979, 1988, 1997, 2006 and 2016.

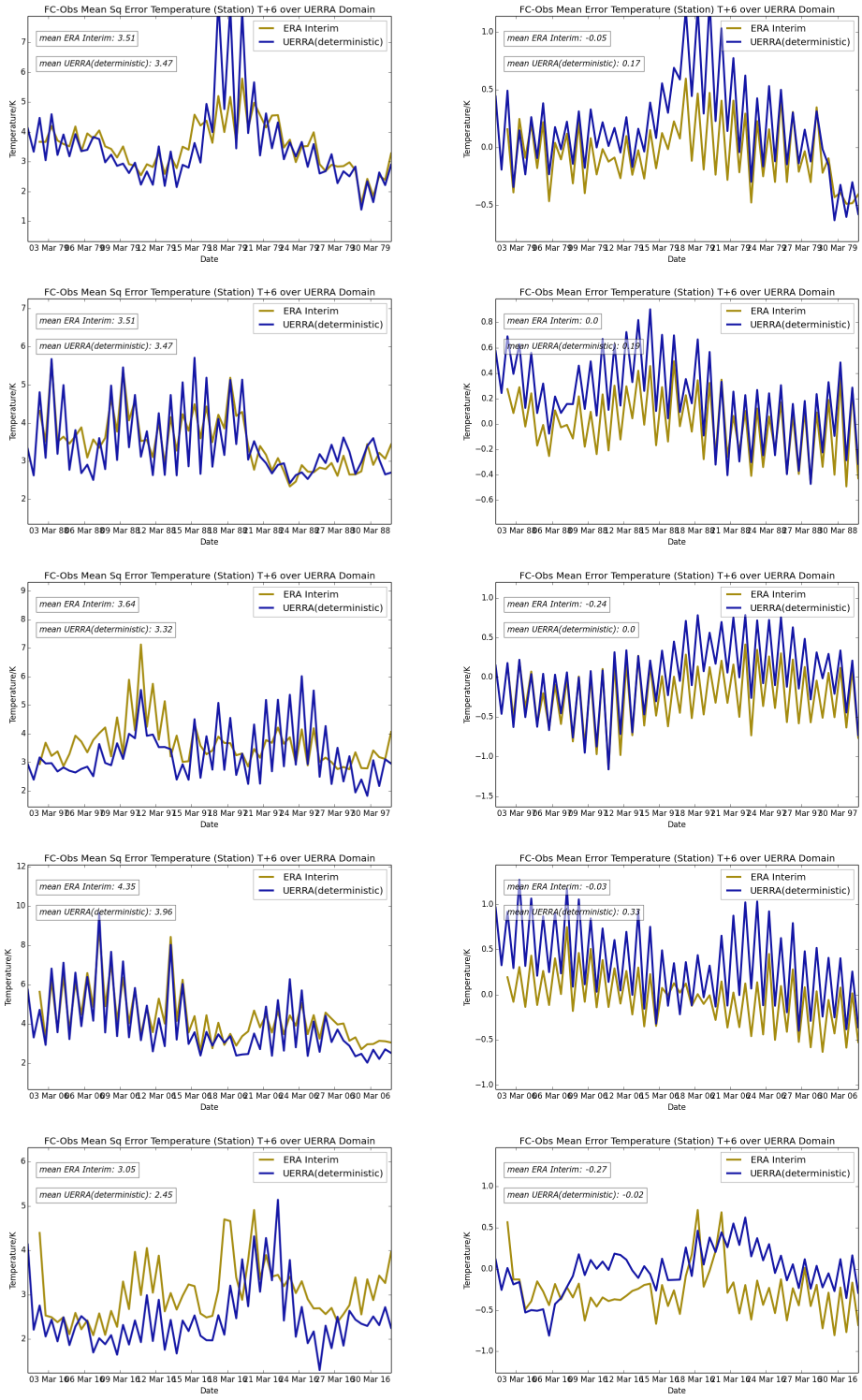

Figure 6: RMSE (LHS) and mean (RHS) difference between 2m temperature observations and a 6h forecast from the deterministic reanalysis. Top to bottom: March 1979, 1988, 1997, 2006 and 2016.

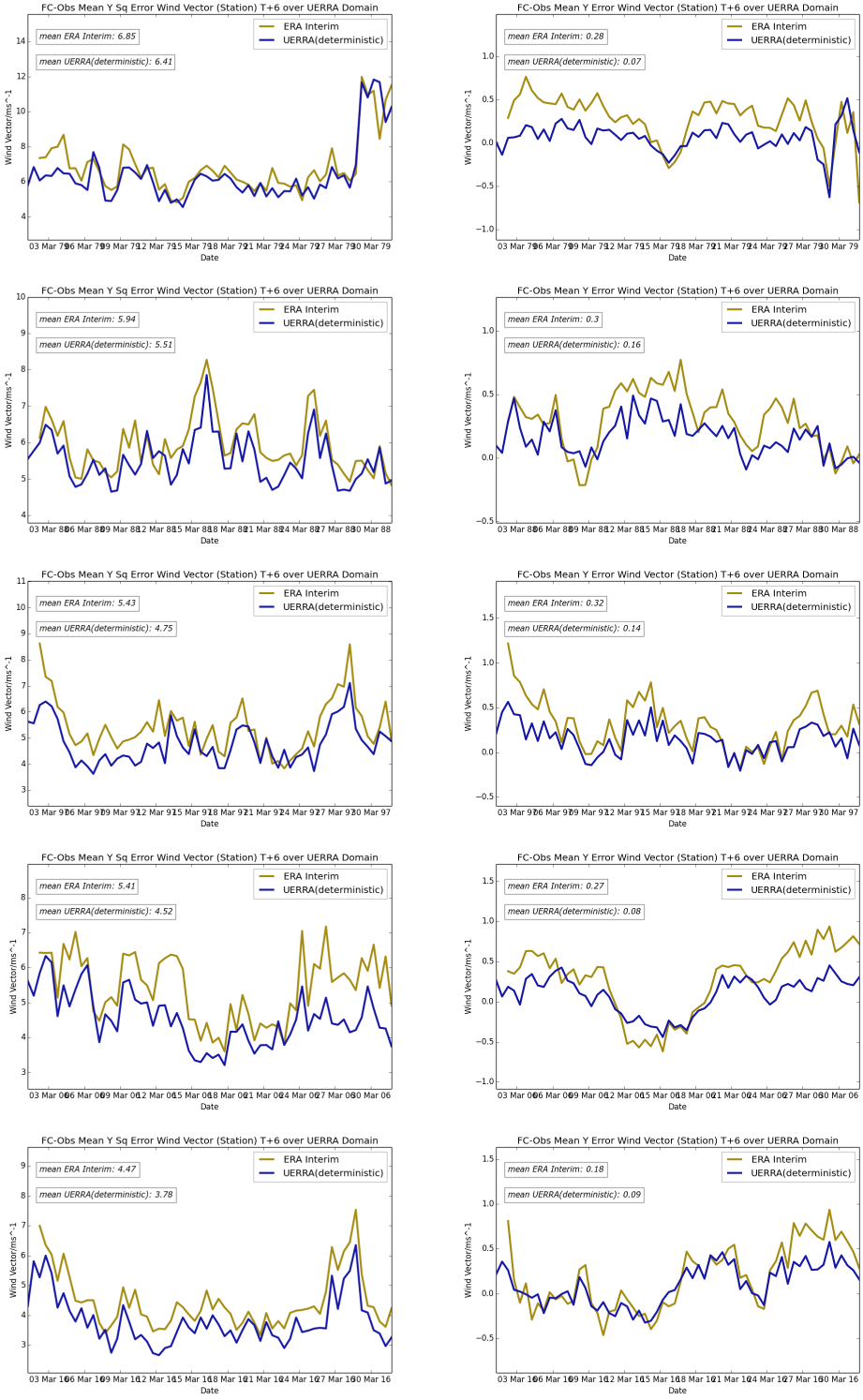

Figure 7: RMSE (LHS) and mean (RHS) difference between 10m wind vector observations and a 6h forecast from the deterministic reanalysis. Top to bottom: March 1979, 1988, 1997, 2006 and 2016.

Year	T2	F10	PMSL	T850	T500	F850	F500
1979	-1%	-6%	10%	-5%	-4%	-3%	3%
1988	-1%	-7%	20%	-8%	-6%	7%	5%
1997	-9%	-10%	4%	-20%	-5%	2%	2%
2006	-9%	-20%	20%	-6%	-10%	5%	10%
2016	-20%	-20%	4%	-30%	0%	4%	-10%

Table 6: Percentage change in RMSE between ERA-Interim and UERRA-MO. Shows 2m temperature (T2), 10m wind vector (F10), mean sea level pressure (PMSL), temperature at 850hPa (T850), temperature at 500hPa (T500), wind vector at 850hPa (F850) and wind vector at 500hPa (F500).

UERRA-MO is an improvement on ERA-Interim throughout the reanalysis period and the improvement increases with increasing time/observation volume, see table 6. Both reanalyses show a small bias, with UERRA-MO consistently less biased than ERA-Interim.

Figure 8 shows the RMSE and mean error of 6h forecasts from the UERRA-MO deterministic reanalysis and ERA-Interim against mean sea level pressure observations for the same periods as figure 6. Again the quality of both reanalyses improves with time/observation volume. However, unlike the other variables, this shows that for pressure UERRA-MO is consistently lower quality than ERA-Interim. This is because pressure is a large scale variable which is better represented by a global model than a regional one. In the earlier half of the reanalysis (1979-1997), UERRA-MO also sees an increase in mean error over ERA-Interim, however in the latter half (2006-2016) UERRA-MO's mean error is a decrease over that of ERA-Interim.

Figure 9 shows the RMSE of 6h forecasts from the UERRA-MO deterministic reanalysis and ERA-Interim against sonde temperature observations for the same periods as figure 6 at 850hPa and at 500hPa. This figure shows similar results to the assessment closer to the surface, with UERRA-MO an improvement on ERA-Interim for each period shown. The improvements with increasing time is not obvious, since the upper air observation network does not increase during the period as significantly as the surface network.

Figure 10 shows the RMSE of 6h forecasts from the UERRA-MO deterministic reanalysis and ERA-Interim against sonde wind vector observations for the same periods as figure 6. These plots show mixed results with eight of the ten periods/heights showing a poorer representation in UERRA-MO than in ERA-Interim. This may be that the winds at these heights are better represented in a global model.

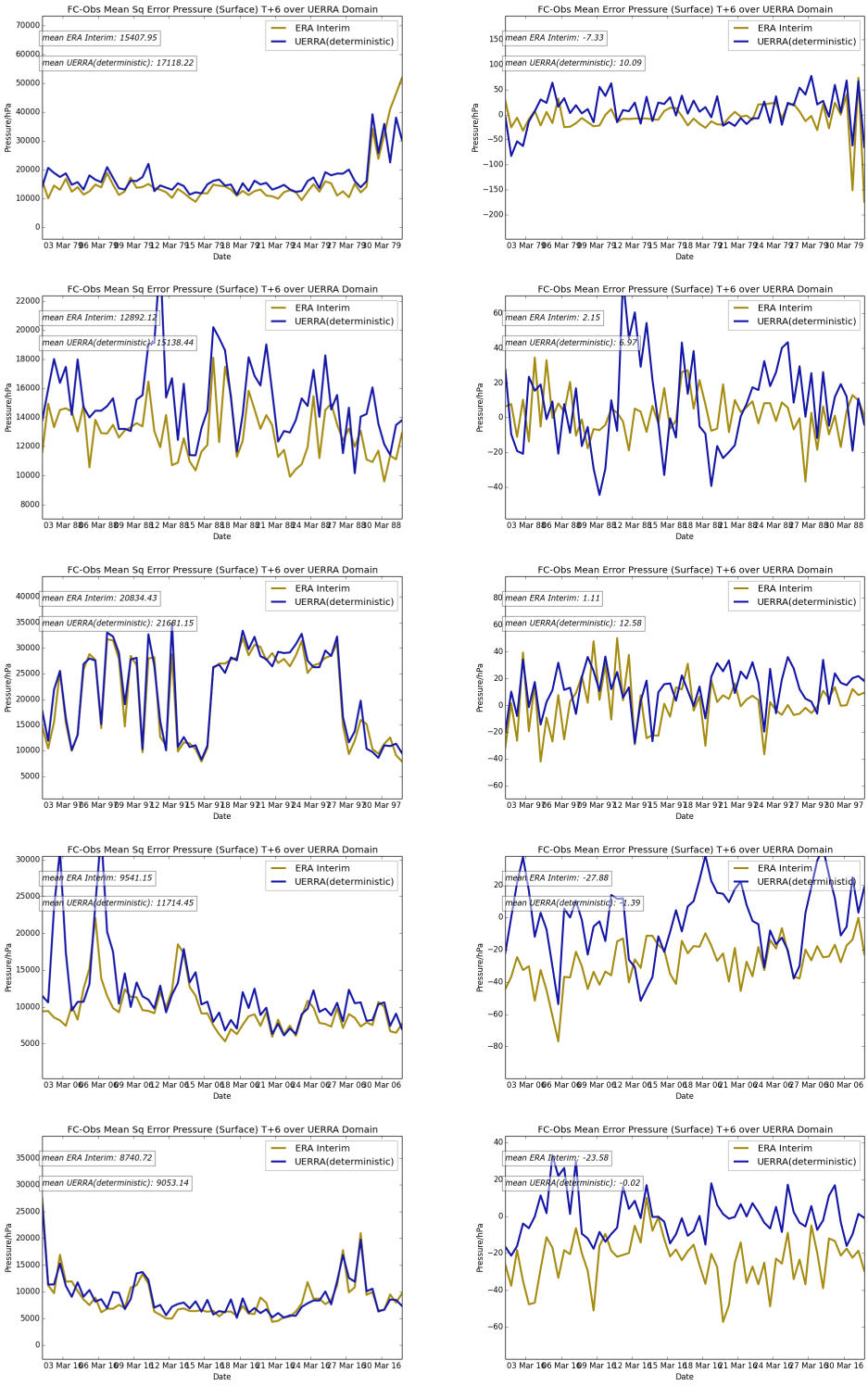

Figure 8: RMSE (LHS) and mean (RHS) difference between mean sea level pressure observations and a 6h forecast from the deterministic reanalysis. Top to bottom: March 1979, 1988, 1997, 2006 and 2016.

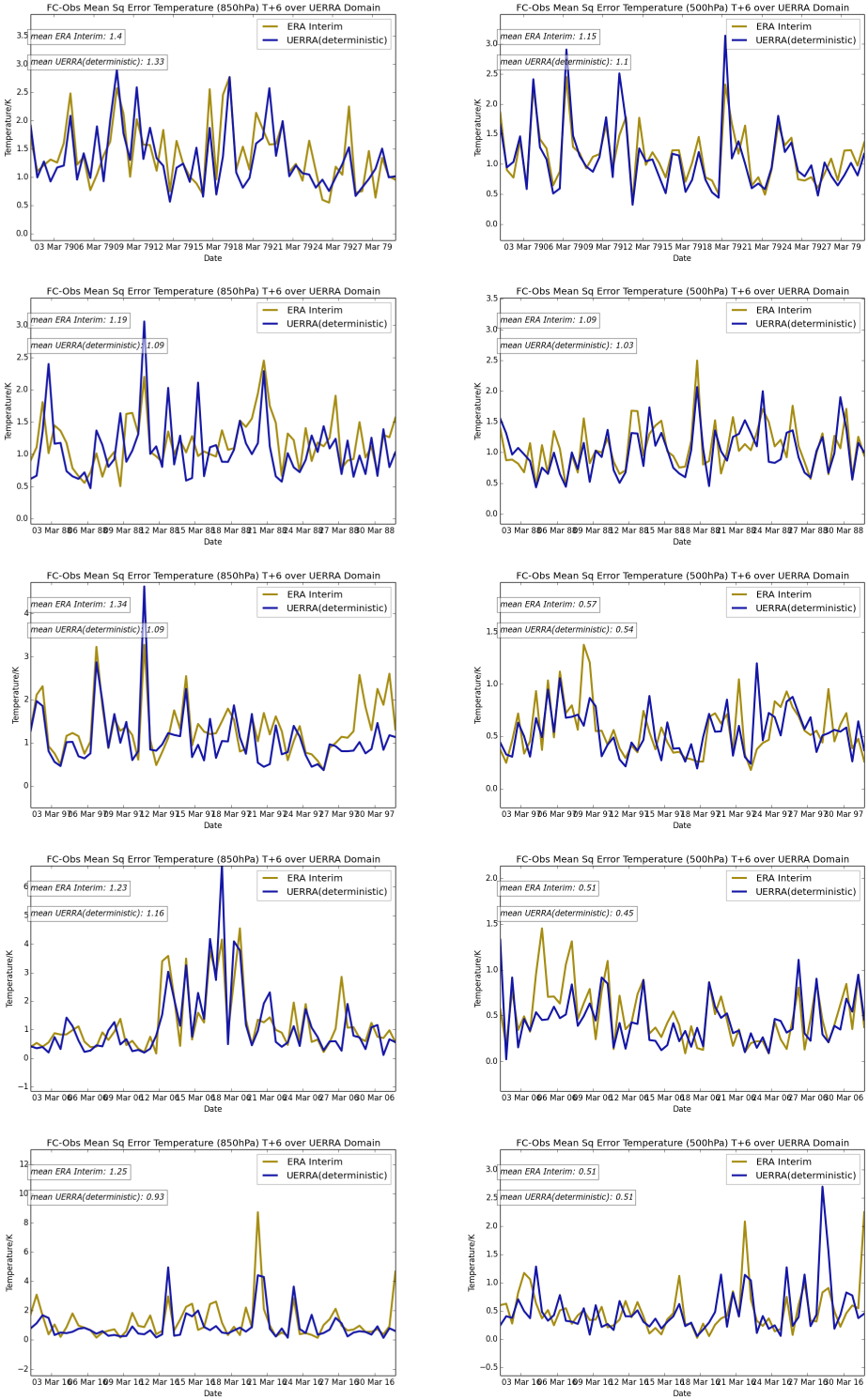

Figure 9: RMSE at 850hPa (LHS) and 500hPa (RHS) difference between temperature sonde observations and a 6h forecast from the deterministic reanalysis. Top to bottom: March 1979, 1988, 1997, 2006 and 2016.

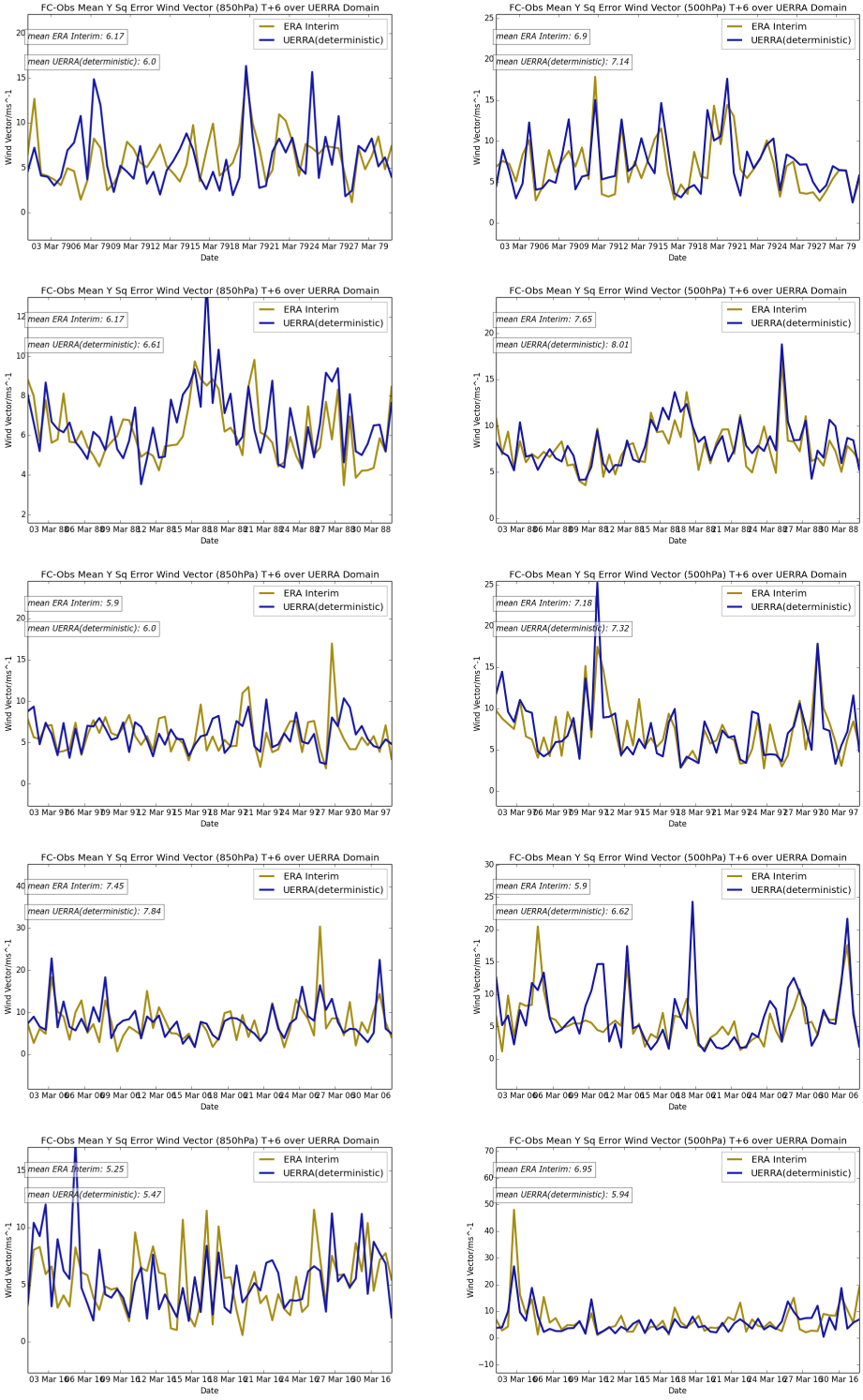

Figure 10: RMSE at 850hPa (LHS) and 500hPa (RHS) difference between wind vector sonde observations and a 6h forecast from the deterministic reanalysis. Top to bottom: March 1979, 1988, 1997, 2006 and 2016.

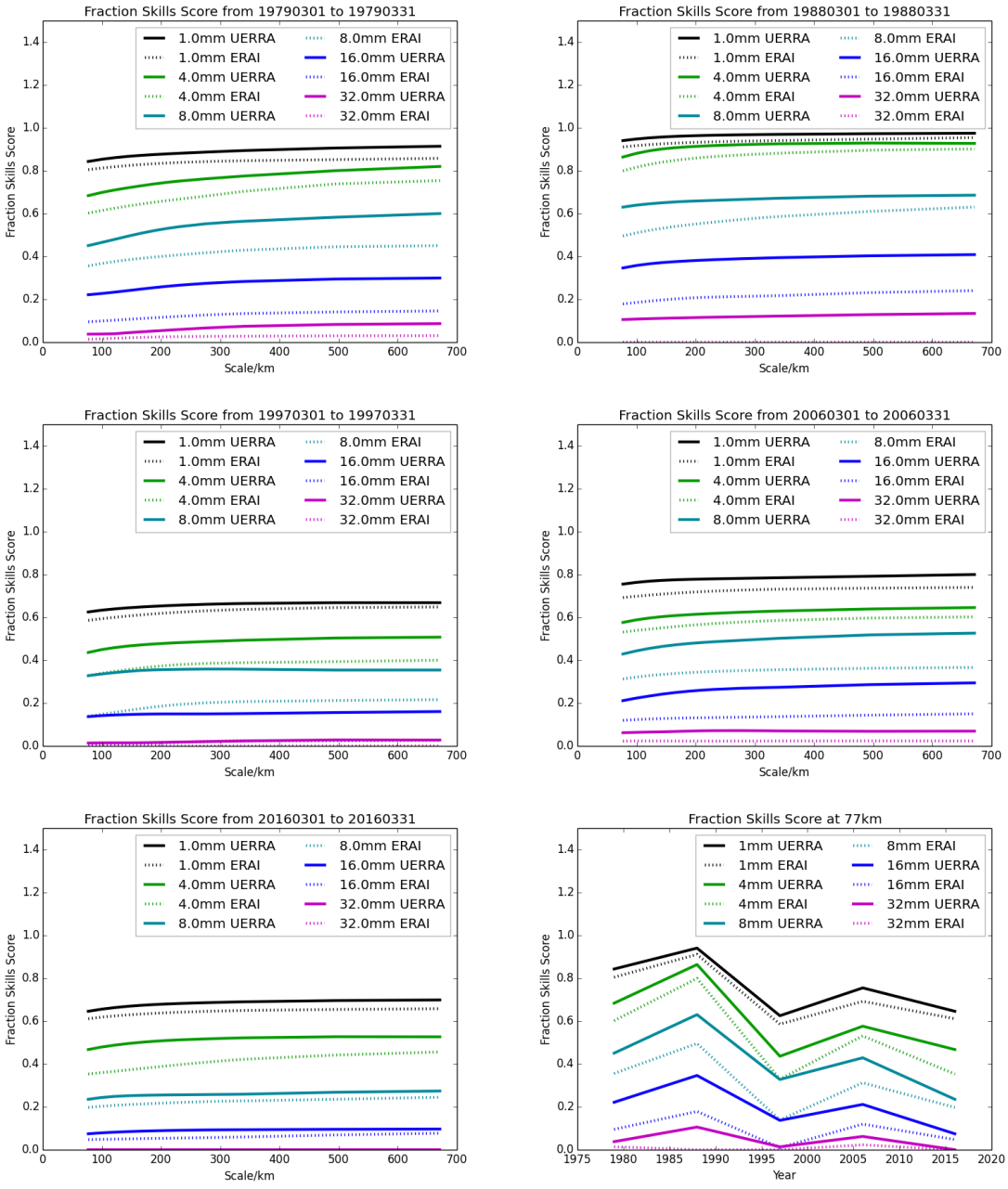

Figure 11: Fractions skill score for 24h precipitation in UERRA-MO and ERA-Interim over Germany, using ECA&D as 'truth'. Top to bottom, left to right: March 1979, 1988, 1997, 2006 and 2016. The bottom right hand plot summarises the others by showing the fractions skill score at 77km against year.

3.3 Fractions Skill Score

The fractions skill score compares the fraction of events in a model with the fraction of events in gridded observations across neighbourhoods of increasing size, [Roberts and Lean, 2008].

Figure 11 shows fractions skill score of 24h precipitation over Germany from the two reanalyses, ERA-Interim and UERRA-MO, using gridded observations ($0.25° \approx 28$km from the European Climate Assessment & Dataset (ECA&D, [Klein Tank et al., 2002]). This demonstrates that UERRA-MO improves on representation of precipitation over ERA-Interim, especially for higher value thresholds.

4 Ensemble Results

4.1 Assimilation Statistics

Figure 12 shows some of the assimilation statistics for the ensemble reanalysis for the month of March from selected years. As in figure 4, the left hand column shows the number of assimilated sondes and the central column shows the number of assimilated temperature observations from surface stations. The ensemble reanalysis assimilates the same observations as the deterministic reanalysis, so again, the number of assimilated surface observations increases greatly over the period of the reanalysis from of the order of 1000 in 1979 to approximately 14000 in 2016 and the number of assimilated sondes remains of the order of 100 at 00Z and 12Z, with fewer at 06Z and 18Z throughout. The occassional drops in the number of assimilated surface observations, seen in figure 12, are also repeated here (Note the different vertical axes).

The right hand column of figure 12 shows the number of iterations of VAR to reach convergence. For this statistic a small amount of variation between members is visible. The decrease in available surface observations towards the end of March 1979 is also visibly reflected in a decrease in iterations for each member.

Figure 13 shows the RMS difference between pressure observations from surface stations and the background (O-B) and between these observations and the reanalysis (O-A) for each ensemble member. Both statistics are larger than their deterministic equivalents, see figure 5. This degradation is primarily due to the decrease in resolution between the deterministic and ensemble reanalyses and the use of 3DVAR in place of hybrid 4DVAR. Both the magnitude and the standard deviation of the RMS differences is reduced moving from O-B to O-A. This indicates that the assimilation draws the reanalyses towards the observations and that the spread of the ensemble increases with forecast time, i.e. it is smallest at reanalysis times. The impact of the change in available observations is manifest in these statistics as small variations in magnitude.

Figure 13 also shows the O-A statistics for potential temperature at model level 5 for each ensemble member. This again demonstrates that there is a small amount of variation between

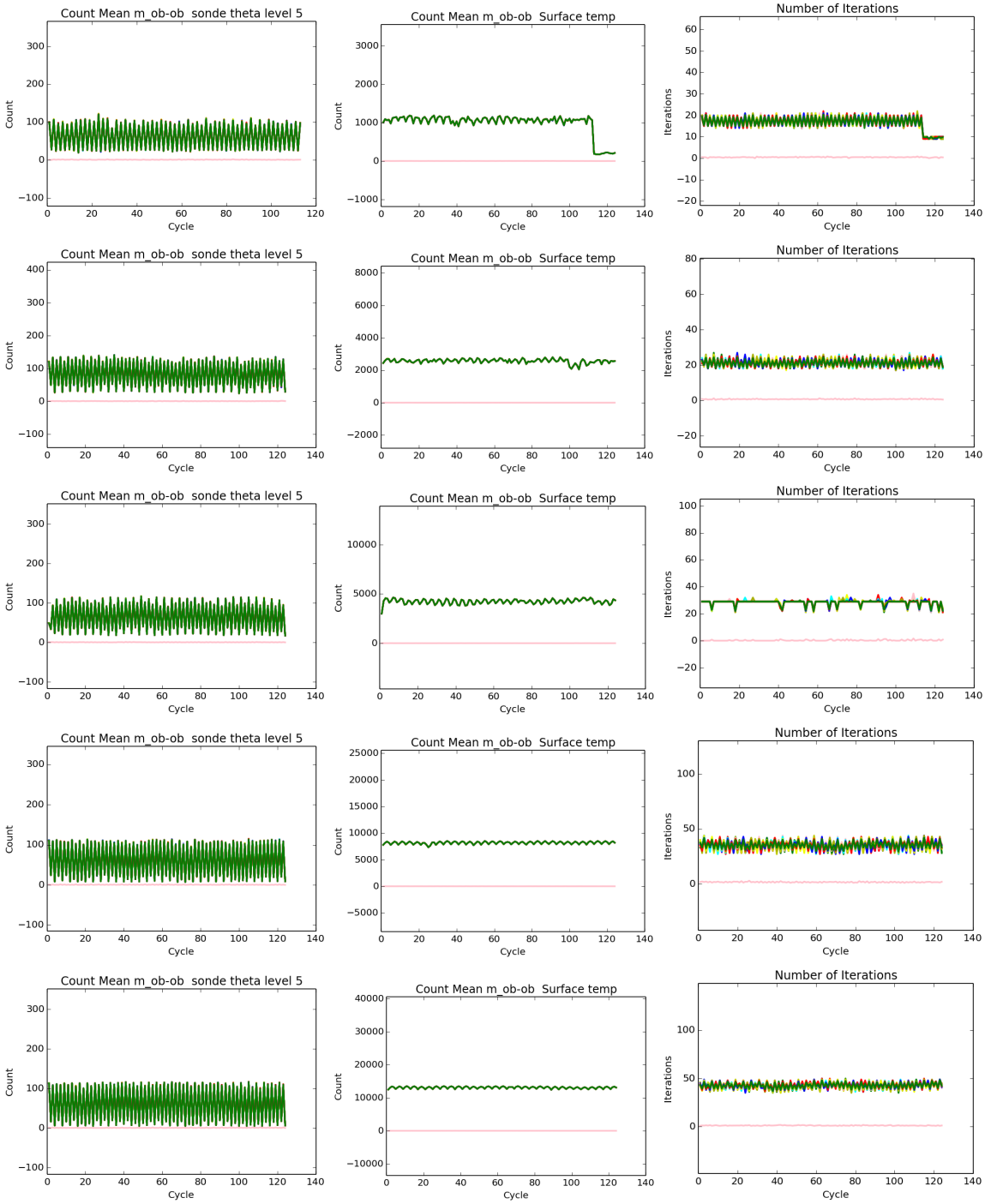

Figure 12: Assimilation statistics. LHS is the number of assimilated sondes. Central is the number of assimilated temperature observations from surface stations. RHS is the number of iterations of VAR to convergence. Top to bottom: March 1979, 1988, 1997, 2006 and 2016. The standard deviation in statistics between the members is given in pink, with values for individual members in other colours. Where only one colour can be seen (green), the difference between the ensemble members is small.

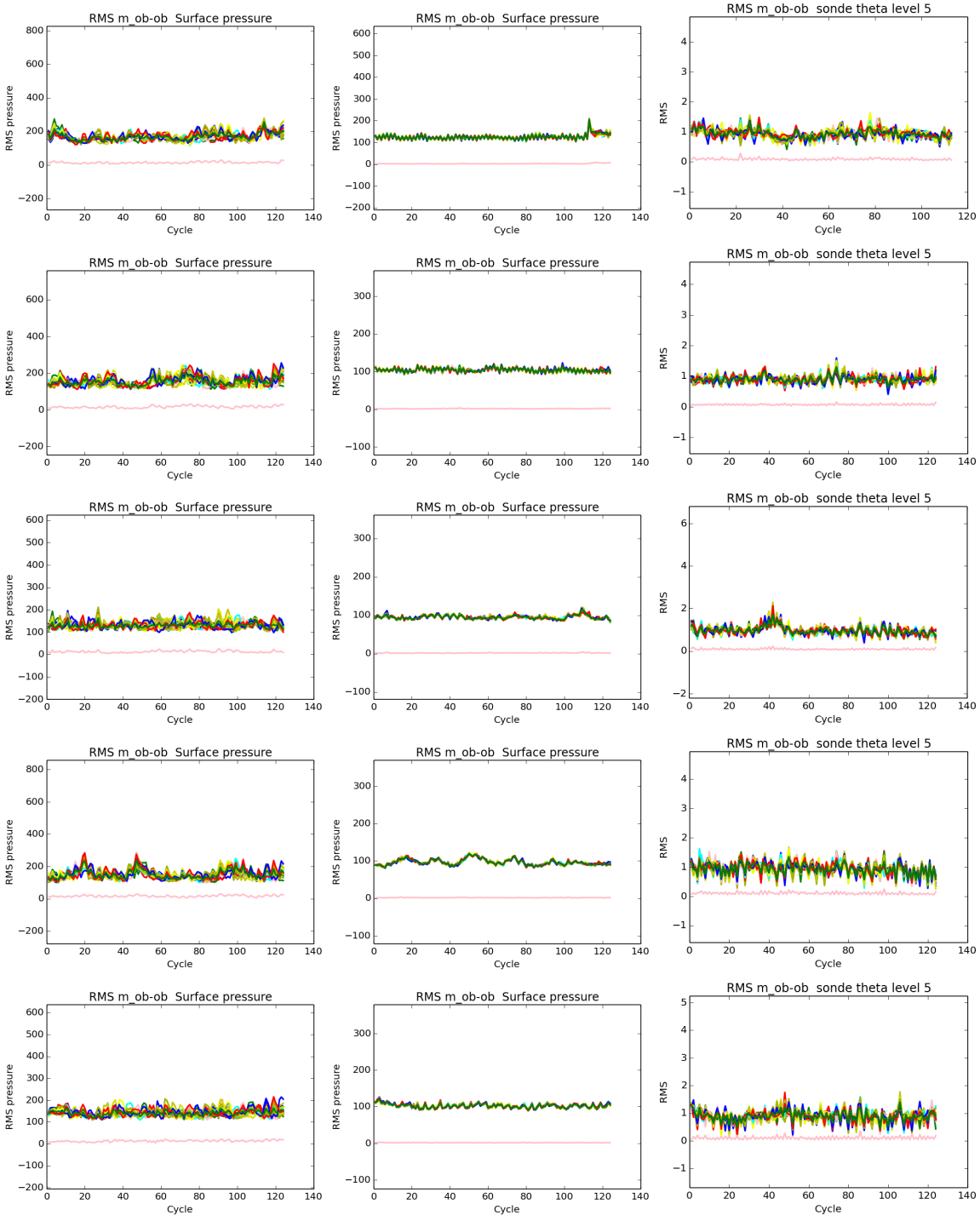

Figure 13: Assimilation statistics. LHS and central are the O-B and O-A, respectively, for SYNOP pressure observations. RHS is the O-A of sonde observations of potential temperature at model level 5 (approximately 980hPa). Top to bottom: March 1979, 1988, 1997, 2006 and 2016. The standard deviation in statistics between the members is given in pink, with values for individual members in other colours. Where only one colour can be seen (green), the difference between the ensemble members is small.

ensemble members, but that the O-A's are reasonably stable throughout the periods shown.

4.2 Ensemble Spread

Ensemble spread is intended to represent the uncertainty in the ensemble mean. Figures 14 and 15 show the spread of each ensemble at observation positions for UERRA-MO and CERA-20C. CERA-20C is a global ten member ensemble reanalysis at 125km and assimilating only surface pressure, marine winds and ocean profiles, [Laloyaux et al., 2016]. The observations are selected daily mean temperatures from the ECA&D dataset, [Klein Tank et al., 2002], which are independent of the reanalyses. The models feature smoothed orography which may differ greatly from individual station height. The model temperatures are first corrected to the height of the observations by applying a lapse rate of $0.0065 Km^{-1}$, [International Organization for Standardization, 1975]. In each case the spread is inflated using a multiplicative factor to match an assumed observation representivity error variance of 0.96 as described in table 7, following [Saetra et al., 2004]. January to March (JFM) is shown for 1987, 1997 and 2007 for both ensemble reanalyses, with 2016 also shown for UERRA-MO (since CERA-20C has no data beyond 2010). These figures show that, after inflating the ensemble to take account of observation representivity error, the magnitude of the spread in UERRA-MO is similar to that of its mean RMSE. However, the same is not true of that of CERA-20C. The Pearson's rank correlation coefficient (labelled 'Prcc'), between the time series of the spread and the RMSE of the mean, is also higher in UERRA-MO than in CERA-20C. This indicates that the ensemble of UERRA-MO is able to represent uncertainties in 2m temperature at least as well as CERA-20C, but at much higher resolution.

Figure 14 also shows that with increasing decade/observation volume, the RMSE of the mean, the spread and the raw spread are all reduced. The correlation of the spread to the ensemble mean is also reduced. This may indicate that the RMSE of the mean is becoming dominated by errors in converting from model space to observation space, e.g. height difference and local effects, or that its variability is becoming dominated by sub-grid scales that are not well captured by simple inflation.

Figures 14 and 15 also show time series of the RMSE of the ensemble mean and the RMSE of the control member, which for CERA-20C is represented by a random member of the ensemble. In each case, the RMSE of the ensemble mean is smaller than that of the control member indicating that for both ensembles the centre of the ensemble is close to the truth.

Figures 16 and 17 show the spread of each ensemble at observation positions for daily mean wind speeds. For the top plots of each figure, 10m daily mean wind speed observations are selected from the ECA&D dataset which are independent of the reanalyses. The model mean wind speeds are not corrected to take account of height differences or any other local feature. Again the spread is inflated using a multiplicative factor, as described in table 8. June to August (JJA) 2007 is shown for both reanalyses. These figures show that the magnitude of the

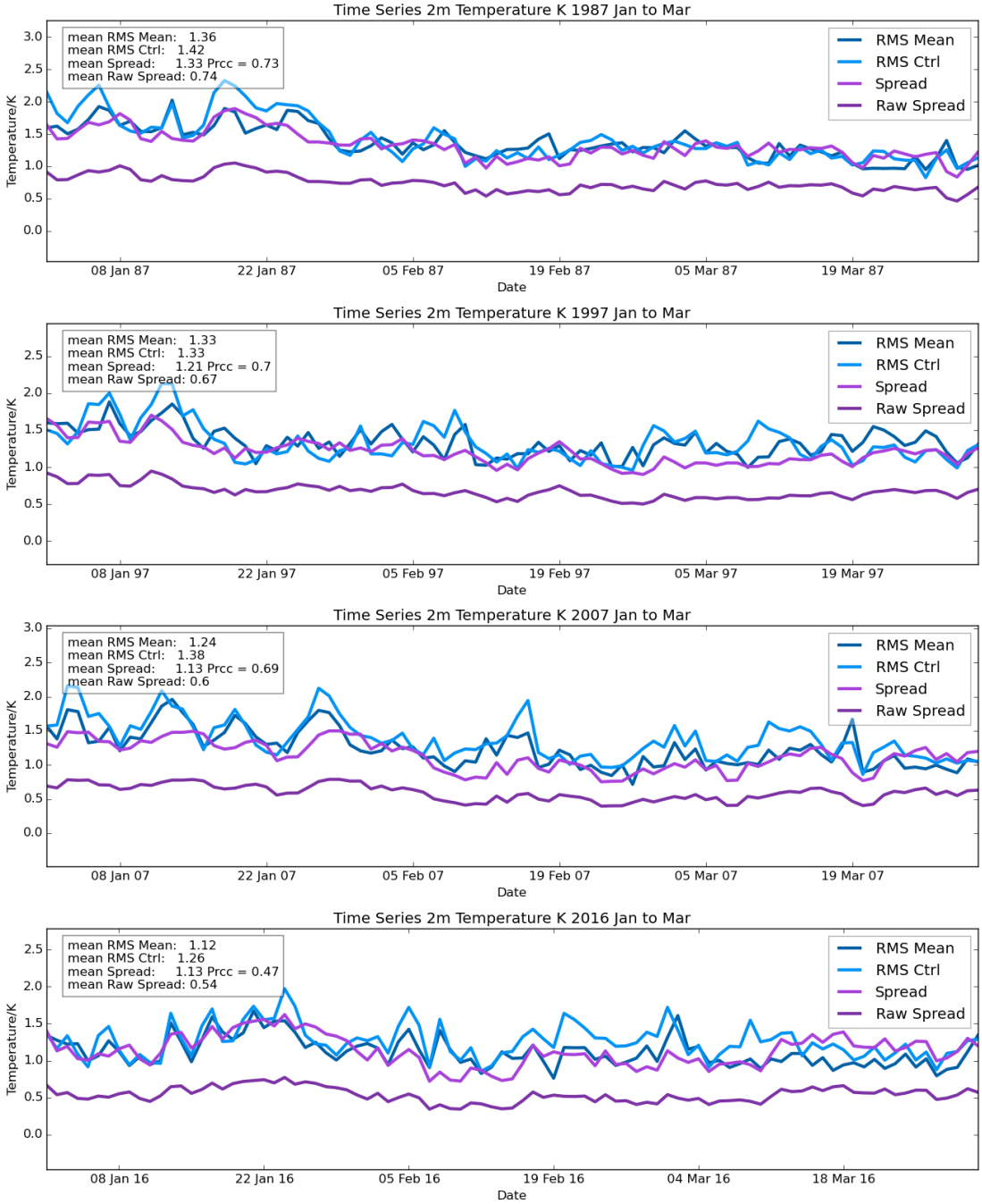

Figure 14: Spread of daily mean 2m temperature in UERRA-MO using ECA&D as 'truth'. Each plot shows raw spread, inflated spread, RMSE of mean and RMSE of control. The Pearson's rank correlation coefficient (Prcc) between the RMSE of the mean and the ensemble spread is also shown. Top to bottom, JFM 1987, JFM 1997, JFM 2007 and JFM 2016.

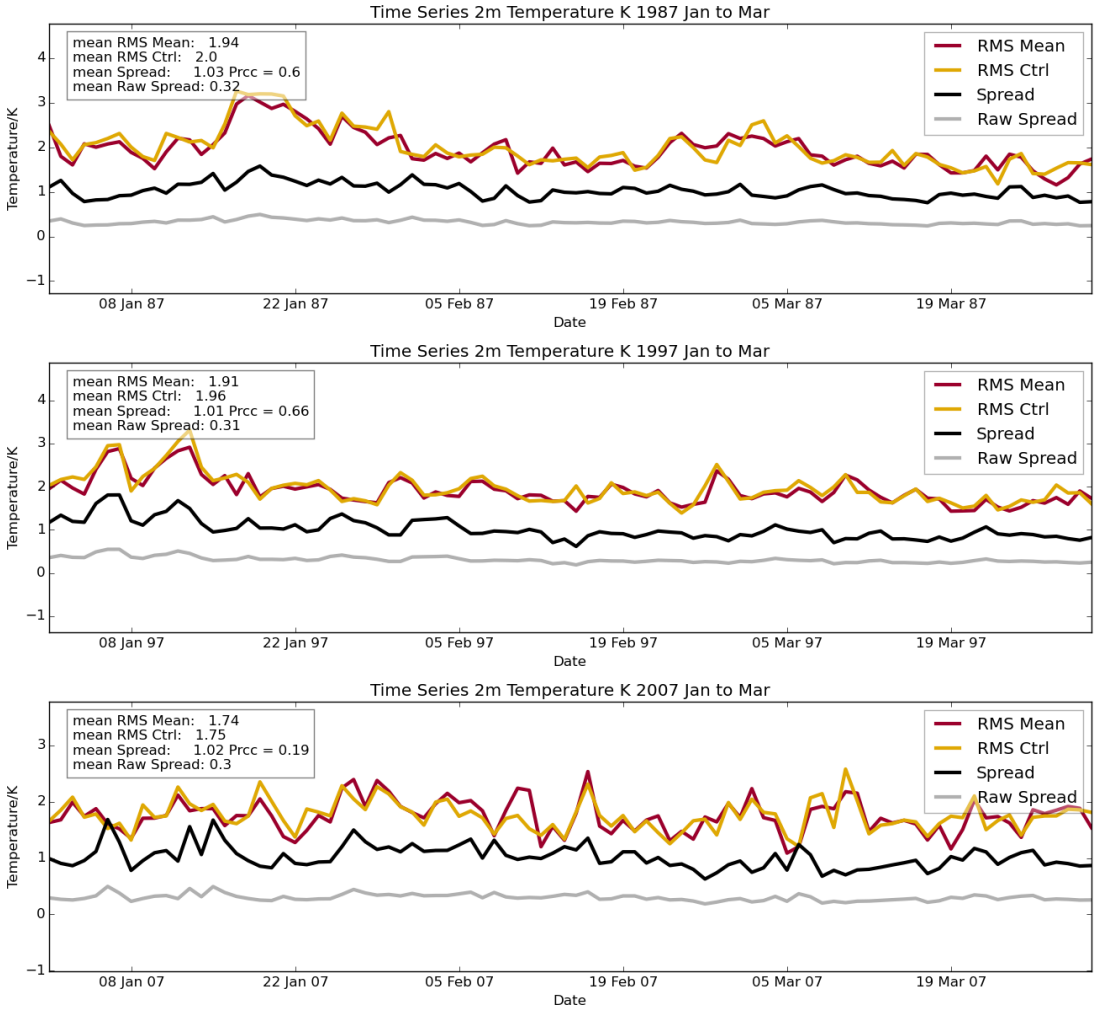

Figure 15: Spread of daily mean 2m temperature in CERA-20C using ECA&D as 'truth'. Each plot shows raw spread, inflated spread, RMSE of mean and RMSE of control. The Pearson's rank correlation coefficient (Prcc) between the RMSE of the mean and the ensemble spread is also shown. Top to bottom, JFM 1987, JFM 1997 and JFM 2007.

Year	Met Office SMR Spread	VR	CERA-20C SMR Spread	VF
1987	0.55	3.2	0.10	9.4
1997	0.45	3.2	0.096	9.9
2007	0.36	3.6	0.090	12
2016	0.29	4.4	n/a	n/a

Table 7: Inflation factors of 2m temperature in UERRA-MO and CERA-20C. In each case the factor is chosen to match an assumed observation representivity error variance of 0.96. Multiplying the Square Mean Raw (SMR) Spread by the Variance Factor (VF) gives the inflated squared spread, which is equivalent to the raw ensemble variance plus observation representivity error variance.

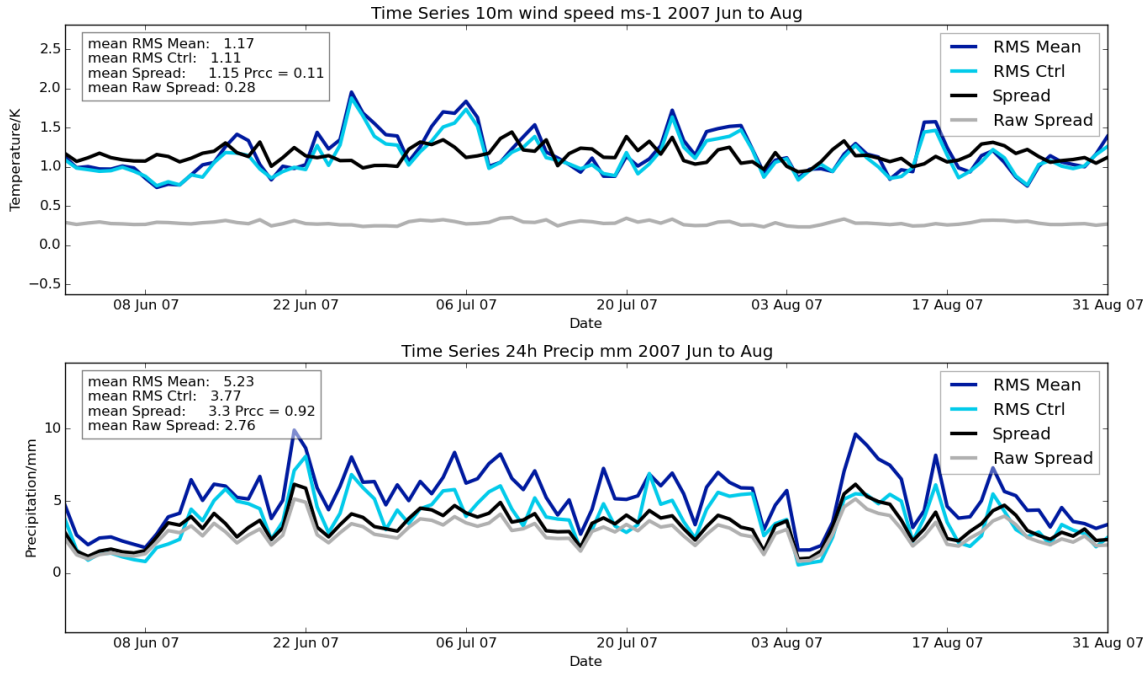

Figure 16: Spread of UERRA-MO using ECA&D as 'truth', JJA 2007. Each plot shows raw spread, inflated spread, RMSE of mean and RMSE of control. The Pearson's rank correlation coefficient (Prcc) between the RMSE of the mean and the ensemble spread is also shown. Top is 10m daily mean wind speed and bottom is 24h precipitation.

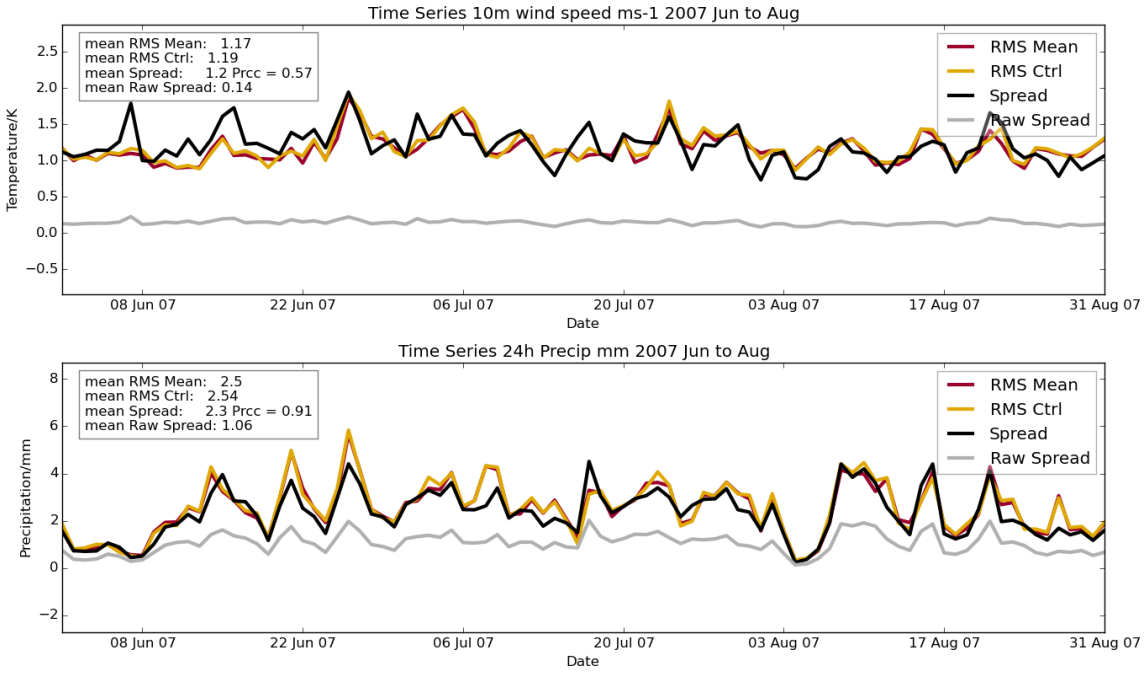

Figure 17: Spread of CERA-20C using ECA&D as 'truth', JJA 2007. Each plot shows raw spread, inflated spread, RMSE of mean and RMSE of control. The Pearson's rank correlation coefficient (Prcc) between the RMSE of the mean and the ensemble spread is also shown. Top is 10m daily mean wind speed and bottom is 24h precipitation.

Variable	Met Office S.M.R. Spread	Spread	V.F	CERA-20C S.M.R Spread	V.F.
10m daily mean wind speed	0.078		18	0.20	77
24h precipitation	7.6		1.4	1.1	2.3

Table 8: Inflation factors of 10m daily mean wind speed and 24h precipitation in UERRA-MO and CERA-20C. Multiplying the Square Mean Raw (SMR) Spread by the Variance Factor (VF) gives the inflated squared spread, which is equivalent to the raw ensemble variance plus observation representivity error variance.

spread is similar to that of the RMSE of the mean for both ensembles, but that the correlation between spread and RMSE of the mean is far greater in CERA-20C than in UERRA-MO. The correlations are lower than those seen for temperature as no local correction is applied, only a simple interpolation.

The bottom plots of figures 16 and 17 show the spread at positions of 24h precipitation observations from the ECA&D dataset, which, again, are independent of the reanalyses. These plots show very large errors in the UERRA-MO ensemble reanalysis. This is due to spuriously large precipitation caused by the shock of a 3DVAR increment into the model. For 3DVAR assimilations it is common not to use precipitation accumulations from the first six hours of the subsequent model forecast, however forecasting beyond the minimum period was prohibitively expensive for this project. Other variables are not affected. This issue is avoided with 4DVAR, as in the UERRA-MO deterministic reanalysis, since these increments are better balanced with the model. In spite of this problem, the spread of the ensemble is very well correlated with the RMSE of the ensemble mean, as is that of CERA-20C.

4.3 Rank Histograms

Figure 18 shows rank histograms of daily mean 2m temperature for the UERRA-MO ensemble for January-March 1987, 1997, 2007 and 2016 and for the CERA-20C ensemble for January-March 1987, 1997 and 2007. These figures suggest that both ensembles are under spread, but that UERRA-MO improves on CERA-20C, perhaps due to the increase in resolution. Additionally, observed values disproportionately are cooler than the ensemble members in UERRA-MO, but warmer than the ensemble members in CERA-20C. This bias reduces with increasing decade for UERRA-MO, but remains large throughout for CERA-20C.

Figure 19 shows rank histogram for daily mean 10m wind speed (top) for the two ensemble reanalyses, JJA 2007. These suggest that the UERRA-MO members more closely match the likelihood of the observed values than those of CERA-20C. The CERA-20C histogram suggests than ensemble members tend to be slower than the observed values. This effect is also present, but reduced in the UERRA-MO histogram.

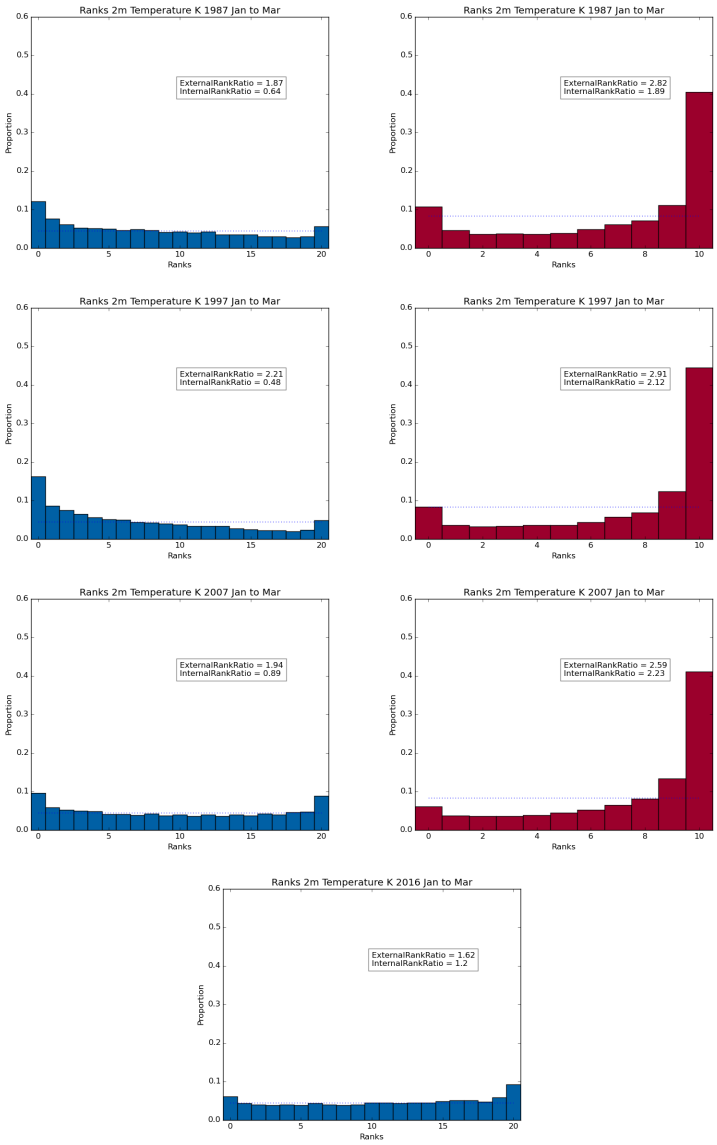

Figure 18: Rank histograms of daily mean 2m temperature in UERRA-MO (lhs) and CERA-20C (rhs) using ECA&D as 'truth'. Top to bottom, JFM 1987, JFM 1997, JFM 2007 and JFM 2016. CERA-20C data is not available for 2016. The external rank ratio is the ratio between the average proportion of the two external ranks to the average proportion of the internal ranks. Perfect spread would result in a score of 1.0. The internal rank ratio is the ratio between the higher half of internal ranks with that of the lower half of internal ranks. An unbiased ensemble would also result in a score of 1.0.

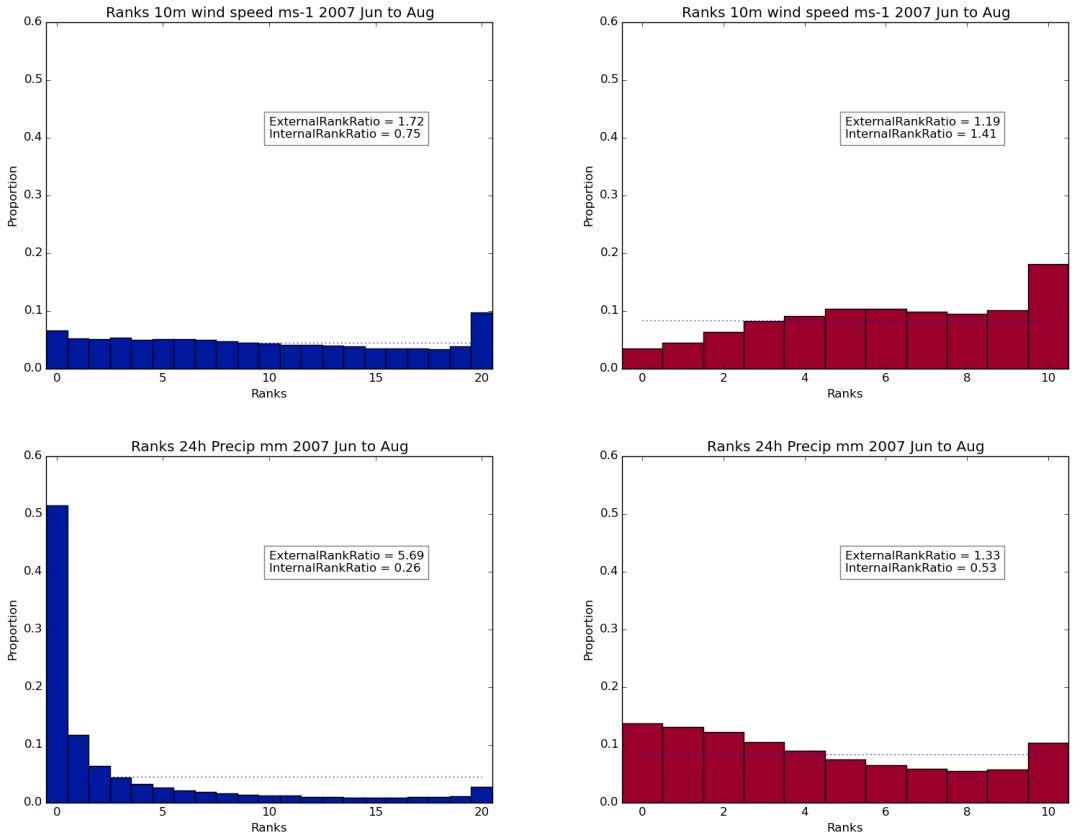

Figure 19: Rank histograms of 10m daily mean wind speed (top) and 24h precipitation (bottom) in UERRA-MO (lhs) and CERA-20C (rhs), during JJA 2007 and using ECA&D as 'truth'. See figure 18.

Figure 19 also shows rank historgram for 24h precipitation (bottom) for the two ensemble reanalyses, JJA 2007. This again shows spuriously large precipitation in the UERRA-MO ensemble which means that members are too often wetter than the observation. A similar effect can be seen to a certain extent in the rank histogram for CERA-20C. In CERA-20C the members tend to be wetter than the observations on most occasions, but there are also too many occassions when the observations are wetter than all of the ensemble members.

4.4 Continuous Rank Probability Scores

The continuous rank probability score (CRPS) is a measure of how well the probability density functions (PDFs) of the ensemble match those of the observations. Time series of CRPS for daily mean 2m temperature in UERRA-MO and CERA-20C are shown in figures 20 and 21 during JFM 1987, JFM 1997, JFM 2007 and JFM 2016 (UERRA-MO only) again comparing against observations from the ECA&D dataset. The scores for UERRA-MO are less than half those of CERA-20C, indicating that UERRA-MO improves in representing the PDFs of the observations. UERRA-MO has twice as many ensemble members as CERA-20C. This gives better sampling of the underlying probability distribution.

Figures 22 and 23 show CRPS time series for daily mean 10m wind speed (top) and 24h precipitation (bottom) for UERRA-MO and CERA-20C, respectively, during JJA 2007. The plots for 10m wind speed show a similar result as that of 2m temperature, with UERRA-MO improving on CERA-20C. CRPS for UERRA-MO precipitation are much higher than those for 2m temperature or 10m wind speed due to the spuriously large values of precipitation in the ensemble. However these scores are still, on average, lower than those of CERA-20C indicating that the precipitation in the UERRA-MO ensemble still contains useful uncertainty information even though its absolute values have large errors.

5 Impact of Satellite Data

The impact of Satellite data on the ensemble reanalysis is demonstrated by comparing the reanalysis with and without Satellite data for a period of three months from three different decades (March to May in the years 1987, 1997, 2007). Two ensemble reanalyses have been produced for each period: one with all available satellite data (Sat, i.e. the production reanalysis) and one without satellite data, i.e. only conventional observations assimilated (NoSat).

As the satellite era has progressed larger volumes of satellite data have become available, as shown in figure 2. The available satellite data for 1987 consists of AMVs from Satwind and the TOVS instrument on a National Oceanic and Atmospheric Administration satellite (NOAA-10). The available satellite data for 1997 consists of Satwind AMVs, scatterometer winds from Scatwind and TOVS from NOAA-12 and NOAA-14. In 2007 significantly more satellite data

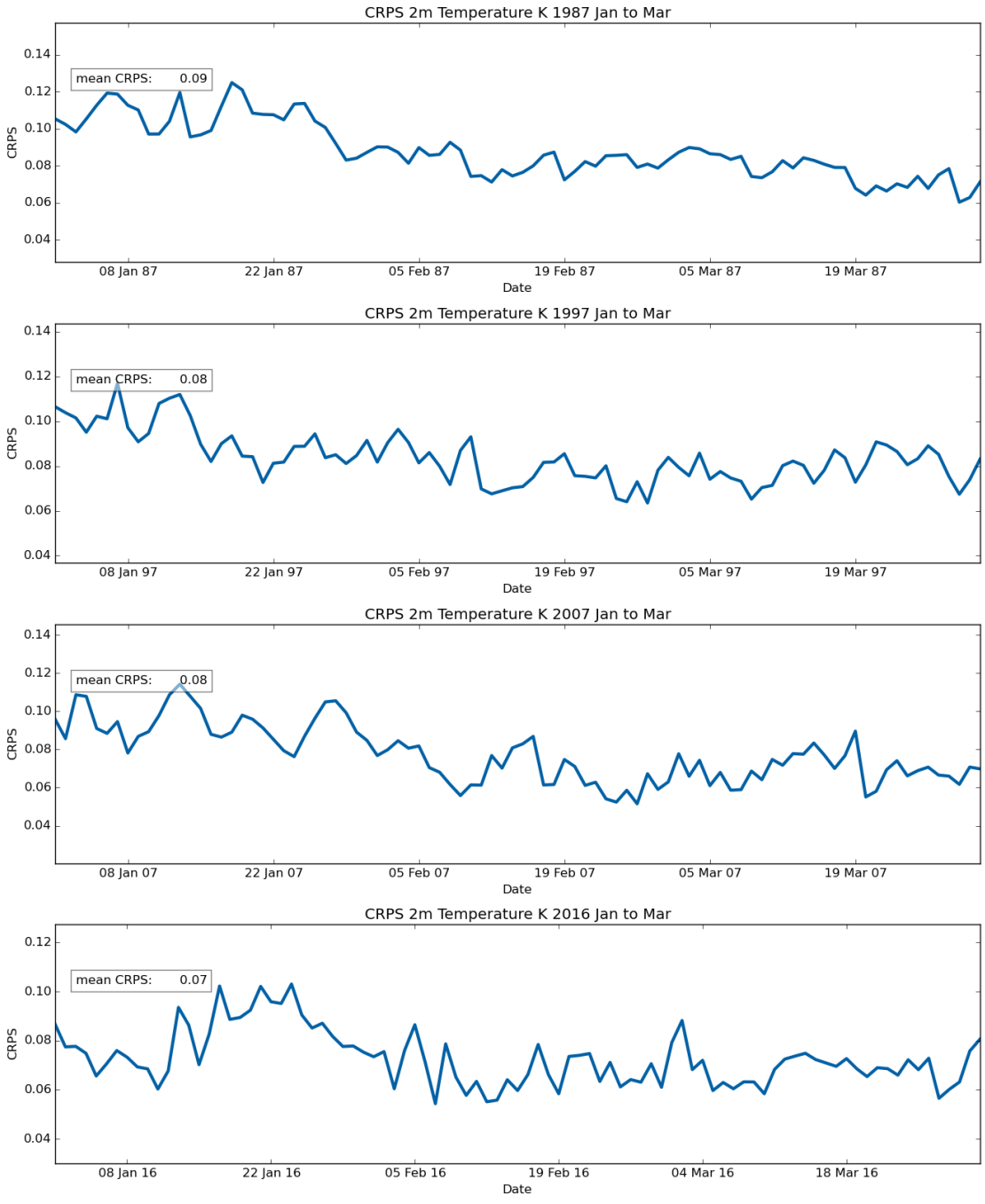

Figure 20: CRPS time series of daily mean 2m temperature in UERRA-MO using ECA&D as 'truth'. Top to bottom, JFM 1987, JFM 1997, JFM 2007 and JFM 2016.

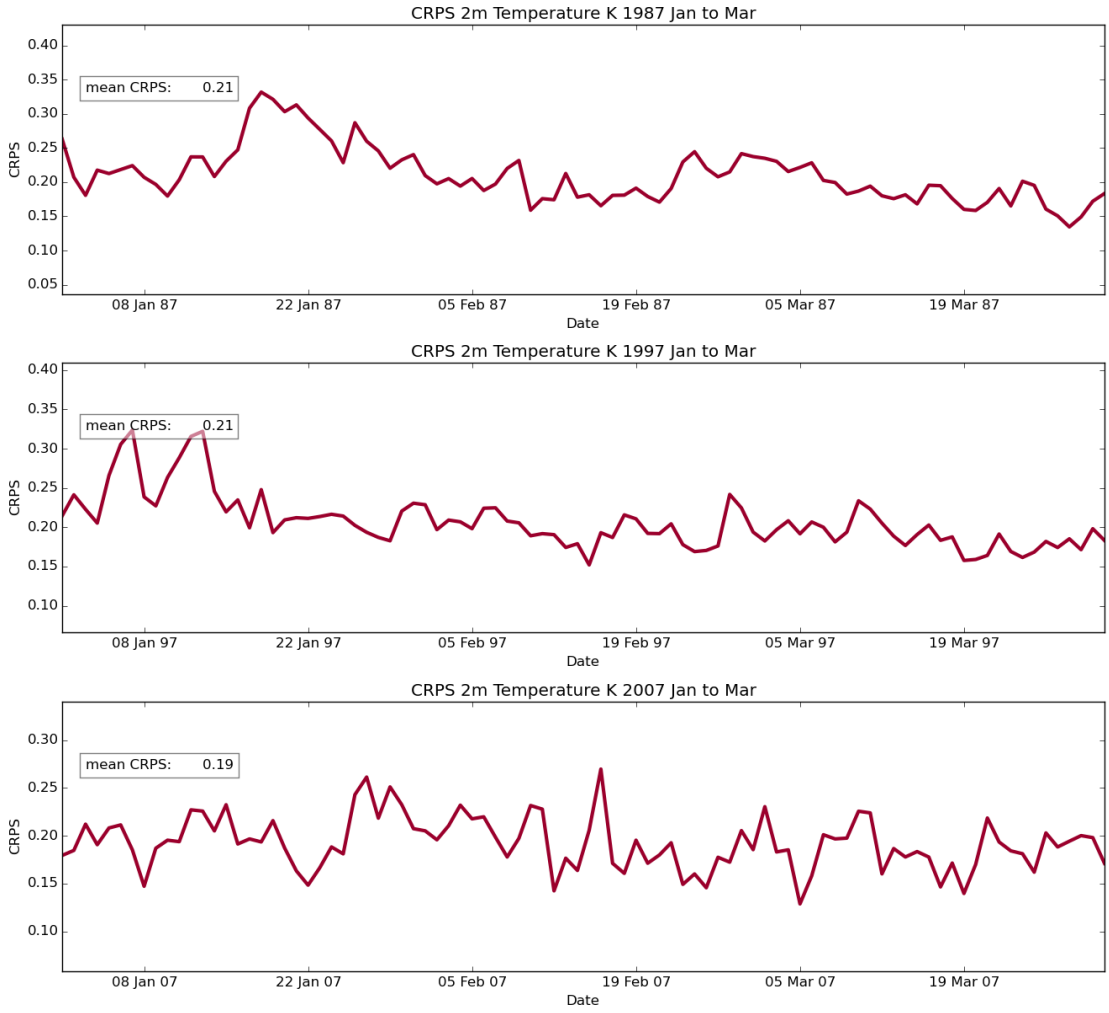

Figure 21: CRPS time series of daily mean 2m temperature in CERA-20C using ECA&D as 'truth'. Top to bottom, JFM 1987, JFM 1997 and JFM 2007.

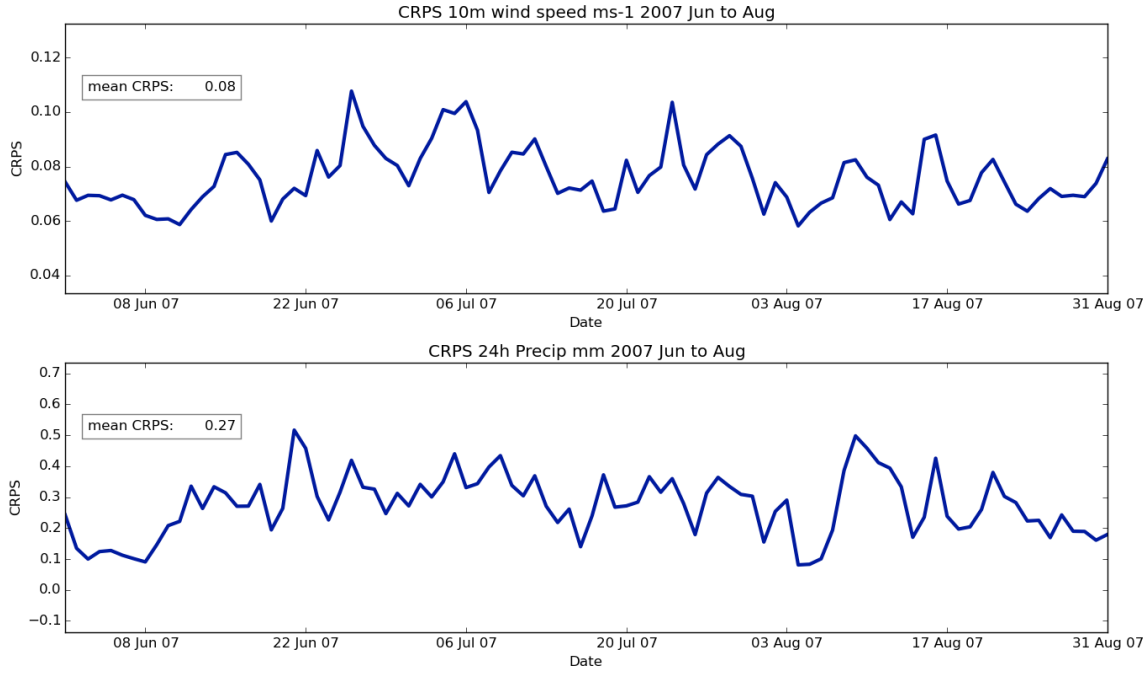

Figure 22: CRPS of UERRA-MO using ECA&D as 'truth', JJA 2007. Top is 10m daily mean wind speed and bottom is 24h precipitation.

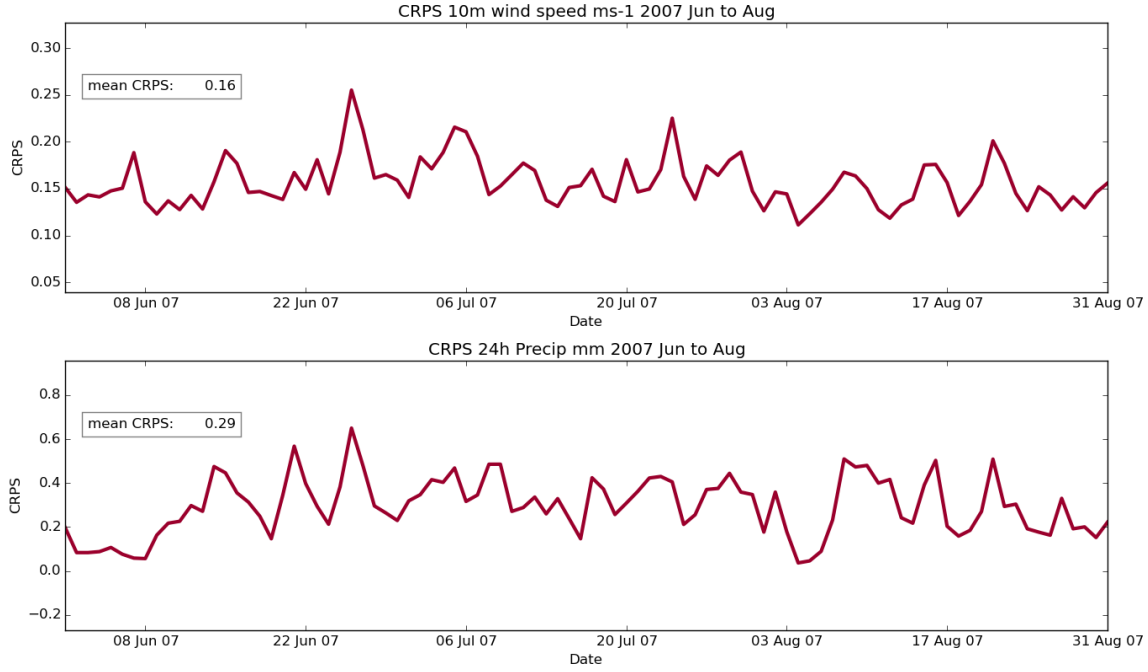

Figure 23: CRPS of CERA-20C using ECA&D as 'truth', JJA 2007. Top is 10m daily mean wind speed and bottom is 24h precipitation.

are available: Satwind AMVs, scatterometer winds from Scatwind, AIRS, and ATOVS from NOAA-15, NOAA-18 and Metop-A from the European Space Agency.

Figure 24 shows the spread and mean error of the ensemble mean (bias) for 250hPa temperature for 1987, 1997 and 2007 for both Sat and NoSat. This shows that the uncertainty is reduced by assimilating satellite data and that this reduction of uncertainty increases with increasing volumes of data. Likewise the bias in the reanalysis mean is also reduced. As expected an increase in the volume of satellite data in 2007 is reflected in the improved performance of Sat over NoSat compared to the earlier periods. This is also seen at other levels (not shown), and reflects the improvement in temperature sounding achieved when TOVS was replaced by ATOVS.

A similar improvement is seen in the 10m wind vector due to the increase in the number of AMVs and scatterometer winds across the three decades. Figure 25 shows the 10m winds mean error of the ensemble mean and spread for each of the three years. Again a reduction in both uncertainty and bias is seen as a result of assimilating satellite data. The impact is largest when most satellite data is available.

6 Archiving

Prior to production archiving, a number of processing calculations are carried out on the atmospheric data. This is because some fields required by the project are not directly available from the model.

- To produce pressure, temperature, wind speed, wind direction, relative humidity, liquid cloud fraction, frozen cloud fraction and cloud cover on levels which represent height above orography, a linear interpolation is performed between model levels. Similarly to produce specific humidity, liquid cloud fraction, frozen cloud fraction and cloud cover on pressure levels, a cubic interpolation is performed between model levels. These interpolations are consistent with those carried out within the atmospheric model.

- The 10m wind direction and wind direction on height levels are calculated by taking the inverse tangent of the true longitudinal component over the true latitudinal component. The true components are calculated by rotating wind components on equatorial latitude-longitude grid, following code in the atmospheric model, see [Mancell and Sharp, 2015].

- To produce geopotential, the geopotential height is multiplied by acceleration due to gravity, approximated by $9.82ms^{-2}$.

- The square root of the sum of the squares of the horizontal wind components is used to calculate wind speed on height levels.

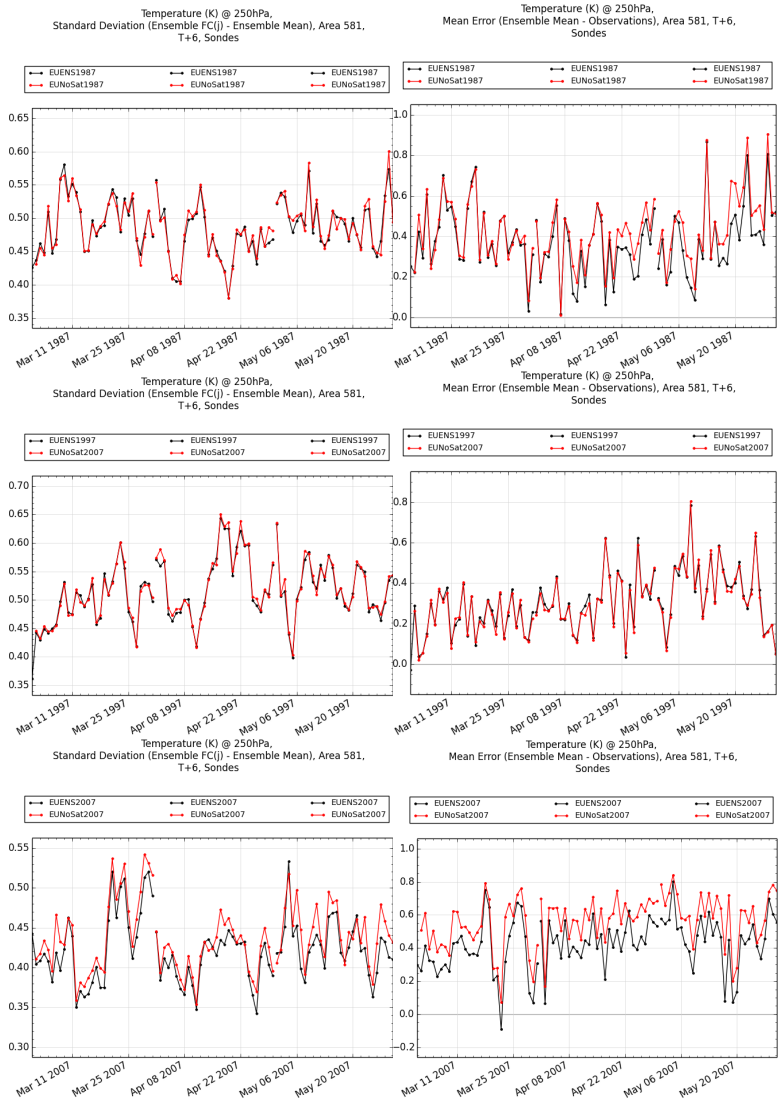

Figure 24: Time series of T+6 forecast data to show the impact of satellite data assimilation. The left hand column shows ensemble spread and the right hand column shows mean error compared to sonde temperature observations at 250hPa. The top row shows March - April 1987, the middle March - April 1997 and the bottom March - April 2007.

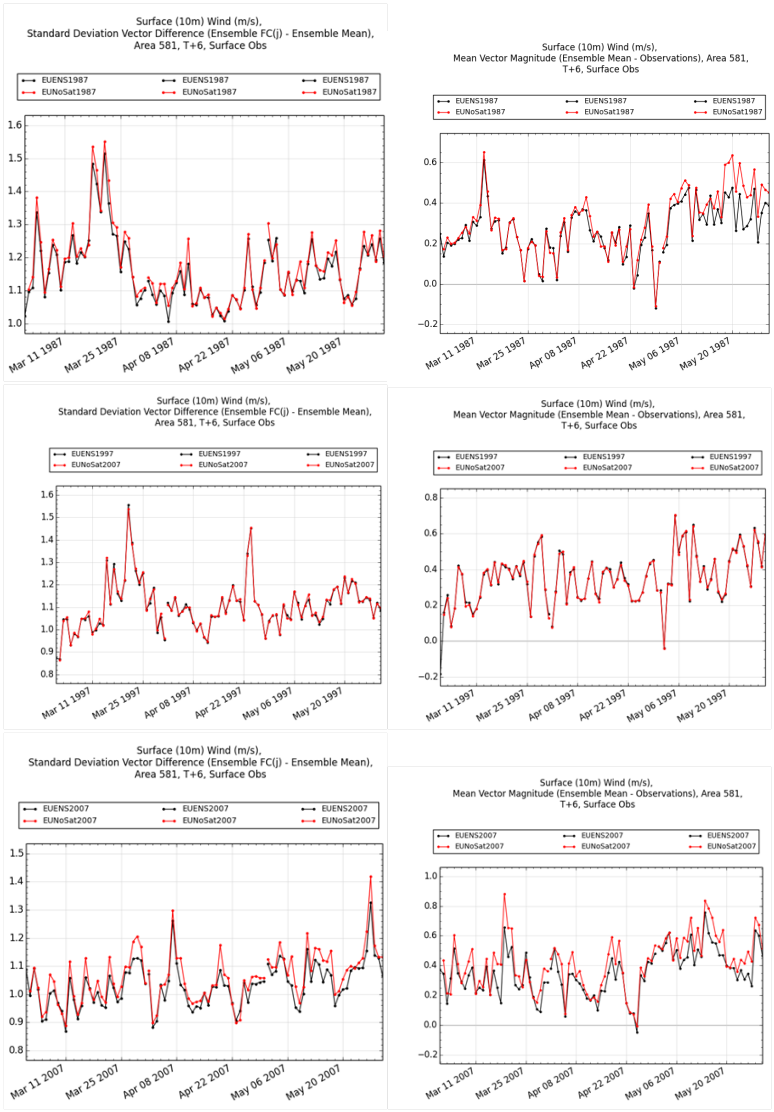

Figure 25: Time series of T+6 forecast data to show the impact of satellite data assimilation. The left hand column shows ensemble spread and the right hand column shows mean error compared to 10m wind observations. The top row shows March - April 1987, the middle March - April 1997 and the bottom March - April 2007.

- Snow fall is calculated as total precipitation minus large scale rain, minus convective rain.
- Total column water vapour is calculated as the total column wet mass minus the total column dry mass.
- Albedo is calculated as the relative difference of net downward short wave radiation with total downward short wave radiation.
- The volumetric soil moisture is calculated by dividing soil moisture by depth and converting from mm into m.

The UERRA-MO reanalyses are stored in ECMWF's MARS archive, [Maass, 2017], under *class=ur, expver=prod, origin=egrr*. The ensemble reanalysis can be selected via *stream=enda* and the deterministic reanalysis via *stream=oper*. For the ensemble reanalysis, (re)analyses are available every six hours - 00UTC, 06UTC, 12UTC and 18UTC, daily, and (re)forecasts at T+1h to T+6h (hourly intervals) from each of these. The deterministic reanalysis is similar, but additionally features (re)forecasts at T+9h to T+30h (three hourly intervals) from 00UTC and 12UTC, daily.

The fields stored in the archive are detailed in appendix A. Model level fields are available at (re)analysis times, accumulated and maximum/minimum fields are avaiable at (re)forecast times and all other fields are available at both (re)analysis times and (re)forecast times.

7 Summary

The UERRA-MO reanalyses cover 1979-2016 over Europe and the Mediterranean. These include a lower resolution ($36km$) 20 member ensemble reanalysis, driven by an ensemble of 3DVAR assimilations, and a higher resolution ($12km$) deterministic reanalysis, driven by hybrid 4DVAR assimilations. The reanalyses assimilate a wide range of conventional and satellite data and a wide range of consistent atmospheric variables is produced from the surface to a height of approximately $40km$.

Both reanalyses make use of technical innovations. The ensemble uses a random draw of analysis increments to represent model error. The deterministic reanalysis uses a hybrid 4DVAR which uses ensemble information to estimate the background error covariances. Both are used for the first time in a regional context.

The parent model for the UERRA-MO reanalyses is ERA-Interim. The deterministic reanalysis shows an improvement of representation of most atmospheric variables over ERA-Interim for this region. This improvement increases with observation volume, indicating that the higher resolution hybrid 4DVAR assimilation system makes better use of a dense network. As expected,

larger scale variables are better represented in the global reanalysis.

Due to cost limitations of running a 40 year ensemble reanalysis, 3DVAR instead of 4DVAR is used to calculate the reanalyses. This restriction has lead to spuriously large precipitation in the initial hours of the (re)forecast. With 3DVAR it is usual to allow a period of spin-up before taking production assimilations. This was not done for UERRA-MO, since extending the (re)forecast period would have increased the cost. Although the subsequent magnitudes of precipitation in the ensemble reanalysis are very large, the ensemble is still able to capture the PDFs of observed values with a similar accuracy to CERA-20C, indicating that there is useful uncertainty information in these fields.

For other fields, the ensemble reanalysis members have realistic representation and are able to capture uncertainties well, demonstrating some improvement on CERA-20C for capturing the uncertainty in the ensemble mean with its spread.

Satellite data has a beneficial impact on the ensemble reanalysis, with later periods where more data is available reaping greater benefits. The effect of the increased volume of satellite data can be clearly seen. The greatest improvement due to satellite data between 1987 and 2007 is seen in the temperature and wind speed, especially in the mid and upper troposphere. This would seem to reflect benefits particularly from the satellite sounding instruments and AMVs. For 10m wind, the verifying observations are mostly over land, while the scatterometer data is from over the oceans and so the impact is harder to measure.

These reanalyses provide the community with hourly data across a wide range of atmospheric variables for 1979-2016. High resolution fields are available from the deterministic reanalysis. The lower resolution ensemble reanalysis provides users with twenty realisations of the entire period whose spread is a useful measure of uncertainty in the mean.

With thanks to

- Rosa Pacione (Agenzia Spaziale Italiana/Centro di Geodesia Spaziale, Italy) and Gemma Halloran for their work in making available reprocessed ground-based GPS data,
- Per Unden and SMHI for providing ECMWF observations in a readable format for 1993,
- The Satellite Applications section of the Met Office for their assistance in gathering and assimilating various sources of satellite data,
- Brett Candy and Breogan Gomez for developing the SURF package to be used in a regional model,
- Adam Clayton for developing hybrid 4DVAR to be used in a regional model,
- Jemma Davie for developing MUSLi,

- Sana Mahmood for setting up our atmospheric model on the ECMWF HPC,
- Adam Maycock for much help decoding observations in ECMWF BUFR format,
- Mike Cullen and Ciara Piccolo for developing a method to perturb the model using analysis increments,
- Marek Wlasak for advice on hybrid covariance settings.

The research leading to these results has received funding from the European Union, Seventh Framework Programme (FP7) under grant agreement no. 607193.

A Catalogue

Available land fields include orography, land-sea mask, surface roughness, snow depth water equivalent, volumetric soil moisture and soil temperature. Volumetric soil moisture and soil temperature are available at four soil levels at depths of 0.1m, 0.35m, 1.0m and 3.0m.

Variable	Surface/Column	Near Surface	H	P	M
Pressure	x	MSL	x		x
Geopotential				x	
Temperature	x	2m	x	x	x
Temperature (Max/Min)		2m			
Wind components (U/V)				x	x
Wind direction		10m	x		
Wind gust		10m			
Wind speed		10m	x		
Relative humidity		2m	x	x	
Specific cloud ice water content			x	x	x
Specific cloud liquid water content			x	x	x
Specific humidity					x
Evaporation	x				
Cloud	x	High/Med./Low	x	x	x
Snow fall	x				
Total Precipitation	x				
Total Column Integrated Water Vapour	x				

Table 9: Atmospheric fields available in the catalogue for both ensemble and deterministic systems. Available height levels (H) are 15m, 30m, 50m, 75m, 100m, 150m, 200m, 250m, 300m, 400m and 500m. Available pressure levels (P) are 1000hPa, 975hPa, 950hPa, 925hPa, 900hPa, 875hPa, 850hPa, 825hPa, 800hPa, 750hPa, 700hPa, 600hPa, 500hPa, 400hPa, 300hPa, 250hPa 200hPa, 150hPa, 100hPa, 70hPa, 50hPa, 30hPa, 20hPa and 10hPa. Model level fields (M, 1-63) are only available at analysis times, all other fields are also available at hourly intervals to a forecast time of 6h. Twice per day (00Z and 12Z) forecast fields from the deterministic reanalysis are available to a forecast time of 27h.

Type	Upwards	Downwards	Net	Direct
LW		x	x	
SW		x	x	x
LW clear sky		x		
SW clear sky	x	x		
Albedo	x			
Latent			x	
Sensible			x	

Table 10: Radiation flux fields available in the catalogue for both ensemble and deterministic systems. Both long wave/thermal (LW) and short wave/solar (SW) radiation are featured.

References

[Bach et al., 2016] Bach, L., Schraff, C., Keller, J. D., and Hense, A. (2016). Towards a probabilistic regional reanalysis system for europe: evaluation of precipitation from experiments. *Tellus A: Dynamic Meteorology and Oceanography*, 68(1):32209.

[Brown et al., 2012] Brown, A., Milton, S., Cullen, M., Golding, B., Mitchell, J., and Shelly, A. (2012). Unified modeling and prediction of weather and climate: A 25-year journey. *Bulletin of the American Meteorological Society*, 93(12):1865–1877.

[Candy, 2014] Candy, B. (2014). Assimilation of satellite data for the land suface. *ECMWF Annual Seminar Proceedings: Use of Satellite Observations in Numerical Weather Prediction*.

[Clayton et al., 2013] Clayton, A. M., Lorenc, A. C., and Barker, D. M. (2013). Operational implementation of a hybrid ensemble/4d-var global data assimilation system at the met office. *Quarterly Journal of the Royal Meteorological Society*, 139(675):1445–1461.

[Davie, 2017] Davie, J. (2017). Monitoring and updating station lists. *varpy*, MUSLi. https://code.metoffice.gov.uk/trac/varpy.

[Davies et al., 2005] Davies, T., Cullen, M. J. P., Malcolm, A. J., Mawson, M. H., Staniforth, A., White, A. A., and Wood, N. (2005). A new dynamical core for the met office's global and regional modelling of the atmosphere. *Quarterly Journal of the Royal Meteorological Society*, 131(608):1759–1782.

[Dee and Uppala, 2009] Dee, D. P. and Uppala, S. (2009). Variational bias correction of satellite radiance data in the era-interim reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 135(644):1830–1841.

[Dee et al., 2011] Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hlm, E. V., Isaksen, L., Kllberg, P., Khler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thpaut, J.-N., and Vitart, F. (2011). The era-interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656):553–597.

[Dee, 2014] Dee, R. (2014). Use of satellite data in reanalysis. *ECMWF Annual Seminar Proceedings: Use of Satellite Observations in Numerical Weather Prediction*.

[Donlon et al., 2012] Donlon, C. J., Martin, M., Stark, J., Roberts-Jones, J., Fiedler, E., and Wimmer, W. (2012). The operational sea surface temperature and sea ice analysis (ostia) system. *Remote Sensing of Environment*, 116(Supplement C):140 – 158. Advanced Along Track Scanning Radiometer(AATSR) Special Issue.

[Gaspari and Cohn, 1999] Gaspari, G. and Cohn, S. E. (1999). Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125(554):723–757.

[International Organization for Standardization, 1975] International Organization for Standardization (1975). Standard atmosphere. *ISO*, 2533.

[Jacob et al., 2014] Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O., Bouwer, L., Braun, A., Colette, A., Dqu, M., Georgievski, G., Georgopoulou, E., Gobiet, A., Menut, L., Nikulin, G., Haensler, A., Hempelmann, N., Jones, C., Keuler, K., Kovats, S., Krner, N., Kotlarski, S., Kriegsmann, A., Martin, E., van Meijgaard, E., Moseley, C., Pfeifer, S., Preuschmann, S., Radermacher, C., Radtke, K., Rechid, D., Rounsevell, M., Samuelsson, P., Somot, S., Soussana, J.-F., Teichmann, C., Valentini, R., Vautard, R., Weber, B., and Yiou, P. (2014). Euro-cordex: new high-resolution climate change projections for european impact research. *Regional Environmental Change*, 14(2):563–578.

[Jermey et al., 2017a] Jermey, P., Bojarova, J., Lockhoff, M., Renshaw, R., Ridal, M., and Unden, P. (2017a). Reanalysis uncertainty evaluation. *UERRA Deliverables*, 2(14).

[Jermey et al., 2015] Jermey, P., Davie, J., Mahmood, S., and Renshaw, R. (2015). Development of ensemble variational data capability and report demonstrating ensemble uncertainty products. *UERRA Deliverables*, 2(1).

[Jermey et al., 2016] Jermey, P., Davie, J., and Renshaw, R. (2016). Ensemble variational da diagnostics. *UERRA Deliverables*, 2(3).

[Jermey et al., 2017b] Jermey, P., Davie, J., and Renshaw, R. (2017b). Ensemble variational da documentation. *UERRA Deliverables*, 2(4).

[Klein Tank et al., 2002] Klein Tank, A. M. G., Wijngaard, J. B., Knnen, G. P., Bhm, R., Demare, G., Gocheva, A., Mileta, M., Pashiardis, S., Hejkrlik, L., Kern-Hansen, C., Heino, R., Bessemoulin, P., Muller-Westermeier, G., Tzanakou, M., Szalai, S., Plsdttir, T., Fitzgerald, D., Rubin, S., Capaldo, M., Maugeri, M., Leitass, A., Bukantis, A., Aberfeld, R., van Engelen, A. F. V., Forland, E., Mietus, M., Coelho, F., Mares, C., Razuvaev, V., Nieplova, E., Cegnar, T., Antonio Lpez, J., Dahlstrm, B., Moberg, A., Kirchhofer, W., Ceylan, A., Pachaliuk, O., Alexander, L. V., and Petrovic, P. (2002). Daily dataset of 20th-century surface air temperature and precipitation series for the european climate assessment. *International Journal of Climatology*, 22(12):1441–1453.

[Laloyaux et al., 2016] Laloyaux, P., Balmaseda, M., Dee, D., Mogensen, K., and Janssen, P. (2016). A coupled data assimilation system for climate reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 142(694):65–78.

[Lorenc, 2012] Lorenc, A. (2012). Variational bias correction of observations. *VAR Scientific Paper*, 32.

[Lorenc et al., 2000] Lorenc, A. C., Ballard, S. P., Bell, R. S., Ingleby, N. B., Andrews, P. L. F., Barker, D. M., Bray, J. R., Clayton, A. M., Dalby, T., Li, D., Payne, T. J., and Saunders, F. W. (2000). The met. office global three-dimensional variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, 126(570):2991–3012.

[Maass, 2017] Maass, C. (2017). Mars user documentation. *ECMWF*. https://software.ecmwf.int/wiki/display/UDOC/MARS+user+documentation.

[Mahmood et al., 2016] Mahmood, S., Renshaw, R., Jermey, P., Doherty, A., Davie, J., and Burrows, C. (2016). Report of observations and datasets assembled for the ensemble-based variational assimilation. *UERRA Deliverables*, 2(2).

[Mancell and Sharp, 2015] Mancell, J. and Sharp, R. (2015). Technical aspects of the reconfiguration. *Unified Model Documentation*, S11. https://code.metoffice.gov.uk/doc/um/vn10.2.

[Oliver, 2017] Oliver, H. (2017). The cylc suite engine user guide. *NIWA*, 7.5.0. https://cylc.github.io/cylc/doc/cylc-user-guide.pdf.

[Pacione et al., 2017] Pacione, R., Araszkiewicz, A., Brockmann, E., and Dousa, J. (2017). Epnrepro2: A reference gnss tropospheric data set over europe. *Atmospheric Measurement Techniques*, 10(5):1689–1705.

[Piccolo and Cullen, 2016] Piccolo, C. and Cullen, M. (2016). Ensemble data assimilation using a unified representation of model error. *Monthly Weather Review*, 144(1):213–224.

[Rawlins et al., 2007] Rawlins, F., Ballard, S. P., Bovis, K. J., Clayton, A. M., Li, D., Inverarity, G. W., Lorenc, A. C., and Payne, T. J. (2007). The met office global four-dimensional variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, 133(623):347–362.

[Roberts and Lean, 2008] Roberts, N. M. and Lean, H. W. (2008). Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Monthly Weather Review*, 136(1):78–97.

[Saetra et al., 2004] Saetra, O., Hersbach, H., Bidlot, J.-R., and Richardson, D. S. (2004). Effects of observation errors on the statistics for ensemble spread and reliability. *Monthly Weather Review*, 132(6):1487–1501.

[Titchner and Rayner, 2014] Titchner, H. A. and Rayner, N. A. (2014). The met office hadley centre sea ice and sea surface temperature data set, version 2: 1. sea ice concentrations. *Journal of Geophysical Research: Atmospheres*, 119(6):2864–2889. 2013JD020316.

[Unden et al., 2014] Unden, P. et al. (2014). Uncertainties in ensembles of regional reanalyses.

[Various, 2017] Various (2017). Rose: A framework for managing and running meteorological suites. *Met Office*, 2017.10.0. http://metomi.github.io/rose/doc/rose.html.

[Wood et al., 2014] Wood, N., Staniforth, A., White, A., Allen, T., Diamantakis, M., Gross, M., Melvin, T., Smith, C., Vosper, S., Zerroukat, M., and Thuburn, J. (2014). An inherently mass-conserving semi-implicit semi-lagrangian discretization of the deep-atmosphere global non-hydrostatic equations. *Quarterly Journal of the Royal Meteorological Society*, 140(682):1505–1520.