Global assimilation of air temperature, humidity, wind and pressure from surface stations: practice and performance

Forecasting Research
Technical Report No: 582

August 2013
Bruce Ingleby

- –

Abstract

Originally the only surface data assimilated in the Met Office global forecasting system were pressure and marine winds but now most temperatures, humidities and winds over land are also used. Adjustments for differences between station and model height are essential for pressure and temperature; new height adjustments for humidity and wind were introduced. These changes brought the global and regional forecasting systems much closer in their use of surface data and forecast performance for surface variables. It was found necessary to exclude winds from islands and headlands not resolved in the forecast model, tropical winds from land stations are also excluded. Extra reports (notably Metars) have been introduced into the system.

The assimilation of land station temperature and humidity reports gave a clear improvement to short range forecasts of "screen" temperature and humidity and small improvements to pressure forecasts. The assimilation of winds over land areas had little impact – wind speed biases, especially at night, are part of the problem. The surface pressure assimilation improves pressure and upper atmosphere forecasts but has little effect on other surface variables. Features of the observation innovations reveal aspects of observation and model errors and other factors such as the proximity to the coast and the importance of the diurnal cycle.

1. Introduction

Near surface temperature and wind are key variables for many users of forecasts. In the past global forecast models have mainly assimilated pressure from surface stations, whilst higher resolution regional models have also assimilated temperature, humidity and wind, plus other variables in some cases. Satellite data have become increasingly important for Numerical Weather Prediction (NWP) but satellite measurements are less useful at low levels over land, due to obscuration by cloud and uncertainties in surface emissivities (despite this satellite-derived land surface temperatures is an area of active research, reviewed by Li et al, 2013). Various changes have been made to the use of surface data in the Met Office global forecast system (Table 1). In 2008 the assimilation of most surface temperature, humidity and wind data was successfully introduced. The short range forecasts of these quantities have to be approximately correct before the data can be successfully assimilated (biases cause particular problems). This has come about due to better representation of soil, boundary layer and cloud properties and higher model resolution.

Date	Change
April 2008	Assimilation of temperature, humidity and wind from most Synops
March 2009	Assimilation of Mobile Synop data (not winds)
March 2010	Improved processing of surface marine data (Ingleby, 2010)
July 2011	Assimilation of Metar data

Table 1. Main changes to the use of surface data in the Met Office global forecast system (see section 2 for observation acronyms). The 2008 and 2011 changes gave moderate impact, the others gave a small impact.

For synoptic scale forecasting in the extratropics surface pressure observations are very useful – the leading mode of forecast errors approximates the barotropic normal mode (eg Ingleby, 2001) and so surface pressure is correlated with pressures and geostrophic winds through much of the atmosphere. Compo et al (2011) used only surface pressure observations to analyse the weather of the entire 20th century. Surface pressure measurement from satellites is discussed by Healy (2013) with a particular emphasis on Global Positioning System radio occultation (GPSRO) measurements. GPSRO data provides some information on surface pressure but this can be affected by forecast temperature biases in the upper atmosphere. Satellite soundings also project onto the barotropic mode (albeit with a need for bias correction). Because the use of surface pressure observations is well established the main focus of this paper is on the global assimilation of surface temperature (T), relative humidity (RH) and wind.

Section 2 describes the surface reports available and reviews their error characteristics. Section 3 introduces the Met Office global NWP system, its observation minus background (*o-b*) statistics (the background is nominally a six-hour forecast) and the changes made. Height adjustment is an important element in making the observation and background as comparable as possible. Section 4 covers the impact of the surface data. Section 5 provides a discussion and summary. An appendix gives details of pressure processing.

2. Observations

Observing standards are coordinated via the World Meteorological Organisation (WMO) and WMO (2010) describes station layout and the various instruments and processing used. Observations are exchanged via the Global Telecommunications System (GTS) in various formats (WMO, 2011). Table 2 summarises the different types of surface observation. The results in this paper are for data in alphanumeric codes (a transition to binary codes – known as BUFR - is underway). For most land stations the station position has to be inserted from a separate database (Section 2.1.4 below). Stations can be manned or partially/completely automated: for pressure, temperature and wind the measurements are essentially the same. Cloud, visibility and current weather are either not available from automatic stations or are measured in rather different ways, humidity tends to be measured differently (Ingleby et al, 2013a). Manned reports are more prone to error in calculation, transcription or transmission (for radiosondes Gandin et al, 1993, documented typical typographical errors: sign error, single digit error or transposed digits). Lazzara et al (2012, Table 2) describe typical problems of Automatic Weather Stations (AWSs) in the Antarctic – similar problems occur elsewhere. Some failures cause loss of data - not readily rectified for remote stations. Occasionally AWSs repeatedly report the same value, but such “stuck values” are quite rare (or are screened out before transmission). Ingleby (2010) found that automated ship reports were slightly better quality than manual ship reports – partly because temperature and humidity sensors can be better exposed (further from the deck) in an automated system, but probably also because of transmission errors in manual reports. Dunn et al (2012) describe various types of error in surface reports.

2.1. Types of report

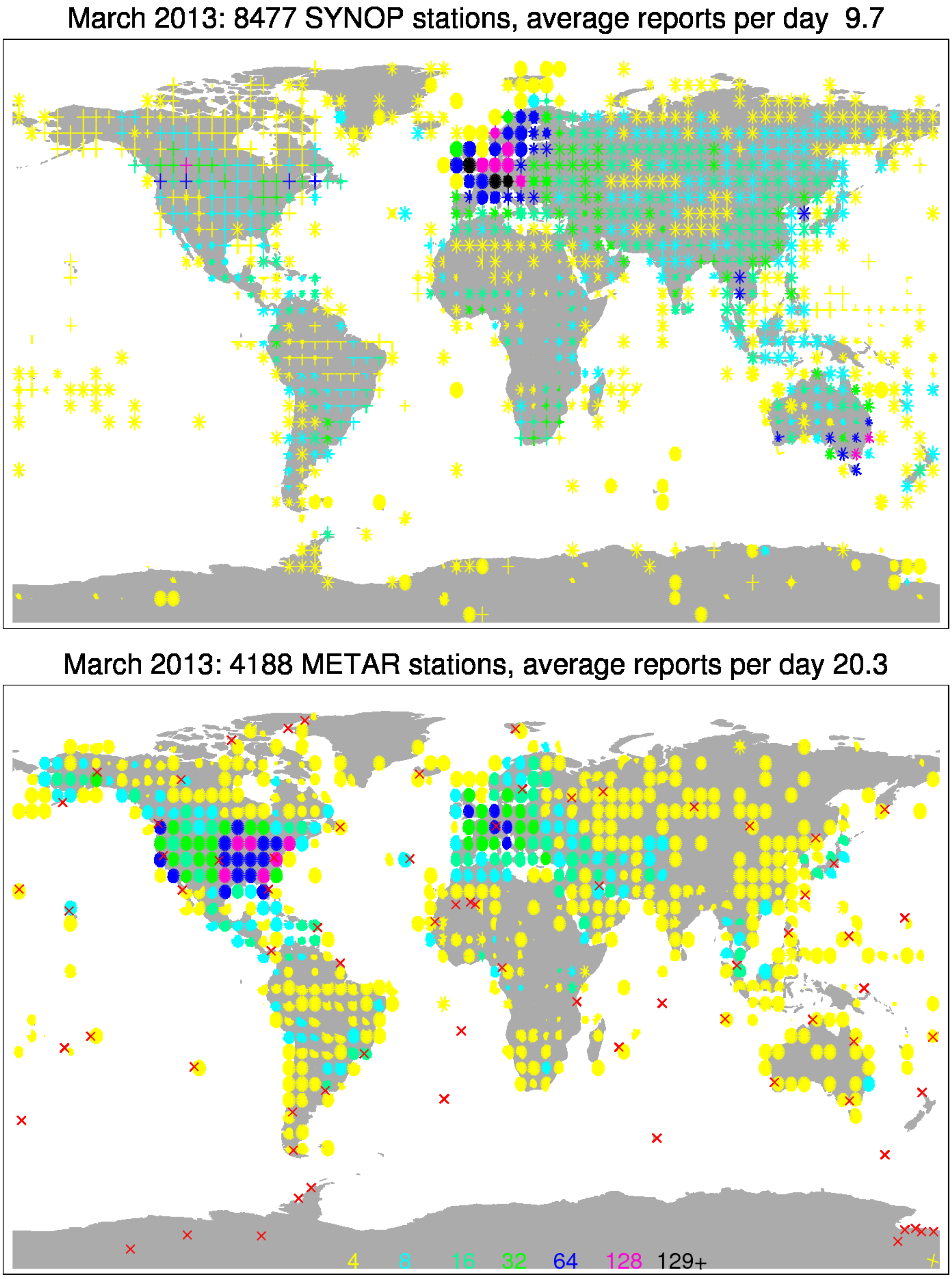


Figure 1. Distribution of Synop (top) and Metar (bottom) reports available at the Met Office in March 2013. The colour coding gives the number of stations within each 5° latitude by 5° longitude box (yellow – up to four stations, light blue – five to eight stations, etc. as indicated by the key at the bottom). For each box the average number of reports per hour (after thinning to one per hour for stations reporting more frequently) per station is shown with 00 UTC reports shown as a spoke to the North and 06 UTC reports as a spoke to the East etc. For reports every hour the spokes merge into a disc, and with 100% availability the discs just touch. On the lower plot red X symbols mark the positions of stations reporting in “Mobile Synop” code.

Name	Nstns Neff	N/6h	Notes	Used
Synop	8500 8100	1-6	A/M, main synoptic network	P,T,RH,W
Metar	4200 2000	6+	A/M, reports from airports	P,T,RH,W
Mobile Synop	100+ 70	6	A, CTBTO, Antarctic, other	P,T,RH
Ship	1900 800	1-6	A/M, some position errors	P,T,RH,W
Moored Buoy	350 330	6	A, coastal, tropical mid-ocean	P,T,RH,W
Drifting Buoy	1000 430	~6	A, mid-ocean, Arctic	P,Arctic T
Platform/Rig	70 70	1-6	A/M, North Sea	P,T,W

Table 2. Summary of surface observations used in Met Office system, see text for acronyms. Nstns is the approximate number of stations reporting in March 2013. Neff is the effective number: the approximate number reporting on any particular day, minus duplicates with Synop for Metar and minus drifting buoys not reporting pressure. N/6h is the approximate number of reports per six hour assimilation window (usually higher for automatic than manned stations; for drifting buoys and remote stations the hours available can be irregular because of the dependence on satellite overpasses). In the Notes section A/M indicates Automated/Manned. Used denotes atmospheric variables assimilated: Pressure, Temperature, Relative Humidity and Wind. See Ingleby (2010) for more details of the marine types.

2.1.1. Synop and Metar

Reports in Synop code form the backbone of the global surface observing network (the number of stations has increased slightly in recent years). The stations are maintained by almost 200 national meteorological services and various agencies working in Antarctica. (Although not on the GTS 125 Met Office automated Climate Data Loggers - reporting temperature and humidity – are treated as Synop reports.) Figure 1a shows that reports from many European stations are available hourly, but for other parts of the world 3- or 6-hourly reports are more usual (efforts are underway to receive hourly data from some additional countries). Station density is highest over Europe and to a lesser extent over parts of Australia, Canada and east Asia. For many stations/regions the report availability (given by the lengths of the spokes) is near 100% but for some (especially in the tropics) the availability is reduced. Almost all stations are land based but some rigs/platforms in the North Sea report in Synop/Metar code (some also report in Ship code).

Metar code is used for reporting hourly (or sub-hourly) weather data at airports, the data volume exceeds that of Synops due to the higher reporting frequency. Almost half of Metar stations are collocated with Synop stations to within 0.02° latitude/longitude. Metar station density is particularly high over North America (Figure 1b; some automatic U.S. Metars have mixed identifiers containing digits - for technical reasons these are not stored in the Met Office data base). Some Metars are restricted for military or other reasons and not circulated freely on the GTS. Synops give temperature and dew point (°C) and pressure (hPa) to one decimal place. Metars report temperature and dew point to the nearest degree. Standard Metars report pressure in whole hPa (rounded down, we add 0.5 hPa to remove the bias caused). About 40% of Metars (notably those from North America) report in hundredths of an inch of Mercury (~0.339 hPa). Despite the lower reporting precision Metar data quality is broadly similar to that of Synops (see section 3.5).

2.1.2. Mobile Synop

The stations using this format (which includes position information) are all fixed and automated, although the Antarctic stations can be relocated occasionally (and some of them move very slowly with the ice cap). There are 60 CTBTO (Comprehensive nuclear

Test Ban Treaty Organization; radionucleide monitoring stations see http://www.ctbto.org/map/#ims) stations, some on remote islands. These have occasional outages of availability. When these stations were first assimilated some missing station heights were obtained from CTBTO and a few were estimated. There are 10 Antarctic stations and also reports from a few stations deployed as part of field experiments: two AWSs in Greenland from late 2009 and FENNEC AWSs (Hobby et al, 2012) in the Sahara from mid-2011. (In 2013 some Indian AWSs started reporting in Mobile Synop code, but each station only reports a few times a month and these are not shown in figure 1).

2.1.3. Other potential sources

In the UK various agencies make roadside temperature and humidity measurements at about 850 locations. There are various sources of surface data on the internet including automated Brazilian stations at http://www.inmet.gov.br/sonabra/maps/automaticas.php and the NOAA mesonet display at https://madis-data.noaa.gov/sfc_display/ which includes some data from the Citizen Weather Observer Program (CWOP) - http://www.wxqa.com/ . Some of the CWOP reports are used in regional NWP, although the winds often have poor exposure and tend not to be used (DiMego, pers. comm. 2009). In 2011 the Met Office supported by the Royal Meteorological Society and Department for Education set up the Weather Observations Website (WOW: http://wow.metoffice.gov.uk/) for the public to submit observations and Bell et al (2013) have performed a preliminary assessment of WOW observations. The CWOP and WOW sources are probably dominated by reports from urban areas. About 80 of the UK roadside reports are now assimilated in the Met Office UK forecasting system (with more to be added soon), but currently none of these sources are used in the Met Office global system. Thus we are gradually moving to a “network of networks” (National Academy of Sciences, 2009). Observations from vehicles (Mahoney and O'Sullivan, 2013) could form one of the networks. From an operational perspective we would want fairly stable sources (in terms of availability and formats) of real-time data – and ideally we want the new data sources to fill gaps in the observation coverage. It is highly desirable to be able to monitor individual stations (sources of data) for biases or other quality issues, it isn’t clear if this would be possible for vehicle data.

2.1.4. Metadata

One of the major issues with surface reports is the availability and correctness of metadata. The Met Office maintains a “Station Master” list with metadata from various sources – updating this is a basic but essential part of using the reports. For Synops the latitude, longitude and station height are available in WMO Publication 9, Vol A (http://www.wmo.int/pages/prog/www/ois/volume-a/vola-home.htm , a high resolution version is available). A small proportion of positions in the list may be erroneous, Ingleby (1995) found height errors in the list. Two heights are given: HP the pressure sensor height and H/HA general station height, or runway height in the case of an airfield. Because pressure is most sensitive to the height Zstn is set to HP if available and to H/HA only if HP is missing. The official International Civil Aviation Organization publication “Location Indicators (Doc 7910)” gives the station names but not their positions! For Metar positions there are only semi-official or unofficial sources – none of these have complete coverage and it is generally unclear how up-to-date the information is (eg http://weather.noaa.gov/tg/site.shtml ,
http://weather.gladstonefamily.net/cgi-bin/wxsite.pl
http://rda.ucar.edu/datasets/ds353.4/inventories/station-list.html).
Almost 3% of Synops are unusable because of unknown positions - sometimes new stations for which the metadata are in transit. 11% of Metar stations are at unknown locations! (Some of the missing Metar positions are being filled in with help from I

Pearmain and A Anglin-Jaffe.) Station identifiers can be reused (as some stations close and others open), and sometimes the old position data can be used by mistake. The new BUFR reports include position information – but experience with radiosonde BUFR reports suggests that this will change the metadata problem rather than eliminate it. (In future when processing BUFR reports we plan to check positions against a separate list and/or against previous reports from the same station.)

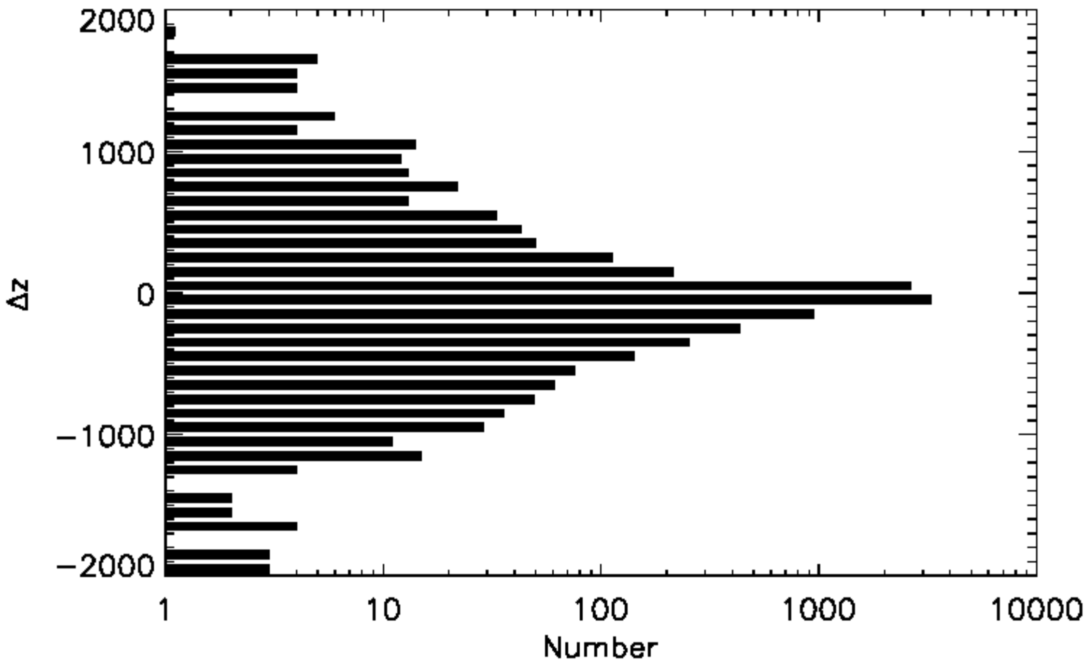


Figure 2. Number of Synop stations (logarithmic scale) with Δz=Zstn-Z_* in 100 m bins, for the Met Office global system in March 2013. There are more stations below the model orography than above it, the mean Δz for this month was -55 m with a standard deviation of 229 m (increasing model resolution slightly tightens the histogram). There were 8515 Synop stations with a height, 41 without; 7393 (8088) had |Δz| less than 250 m (500 m). For Metars (not shown) the mean Δz was -52 m with a standard deviation of 175 m.

Synop stations range in height from -350 to 4107 m above sea level. As seen in Figure 2 there is a tendency for observing stations to be in lowland areas and hence below the model height (Z_*, this is a grid box mean height, bilinearly interpolated to the station position). There are a few mountain top stations with very large Δz=Zstn- Z_* values. Some of the highest stations in the world are actually valley sites in Tibet (Pepin and Seidel, 2005). Most of the largest |Δz| values occur in the Alps or Himalayas.

2.2. Data measurement and errors

2.2.1. Pressure

All Synop reports should include the pressure as measured (Pstn) and either pressure adjusted to mean sea level (Pmsl) or, for high level stations, the height of the 850, 700 or 500 hPa surface. Pmsl becomes increasingly ill-defined for higher stations (different countries use different adjustment algorithms) and in principle the use of Pstn is much cleaner, although in some cases it is compromised by erroneous Zstn values (Ingleby, 1995). Some stations show pressure biases, but it can be difficult to distinguish between calibration and height errors. At the Met Office both Pstn and Pmsl are monitored on a monthly basis and either of them can be converted to P_* (pressure at model height) for use in the assimilation. Metars report pressure at sea level calculated using a standard atmospheric profile – we invert this to give Pstn. See Appendix 1 for details of pressure processing.

2.2.2. Temperature

Temperature and humidity are measured in screens between 1.25 and 2 m above the ground. (Snow cover can change the effective screen height and some countries adjust the screen height accordingly – but presumably not at remote automatic stations.) Temperatures measured in standard screens are subject to overheating in direct insolation under weak wind conditions (and more minor cooling problems on calm, clear nights). Such overheating can be alleviated using aspirated measurements (a fan pulling air over the instruments – this needs careful design or the fan can cause problems), these are used in the USA and Japan but probably not many other countries because of the increased cost and power consumption. The overheating can be up to 1 or 2 degrees for a few hours around noon (see Painter, 1977, Hubbard et al 2004 and Nakamura and Mahrt, 2005; Lin and Hubbard, 2008, also discuss lag effects). The overheating problem is worse over snow (and possibly sand) as the surface reflects additional sunlight into the screen. Arck and Scherer (2001) and Lin et al (2005) suggest this can give an additional 2 or 3 degree error in extreme conditions (so 5 or 6 degrees in total). Genthon et al (2011) report that biases occasionally exceed 10 degrees on the Antarctic plateau. To correct such radiation biases would be difficult (and depends on the screen used); it is more feasible to reject the temperatures likely to be worst affected (not current practice at the Met Office).

The diurnal temperature range depends on soil moisture and cloud cover. Locally surface temperature and wind speed may be positively correlated, at least during the diurnal cycle (Lapworth 2003). Local temperature and humidity perturbations (and forecast errors, typically either warm and dry or cool and moist) tend to be negatively correlated over land – related to the partitioning of surface fluxes between sensible and latent heat. Surface temperature has a very wide range globally, with a slight peak in the frequency near 0°C corresponding to melting snow or ice.

2.2.3. Humidity

Humidity is usually measured in the same screen as temperature. Traditionally psychrometers (wet and dry bulb thermometers) were used but, partly as a result of automation, many stations have changed – usually to capacitive sensors. Ingleby et al (2013a) provides a description of the instruments and results from field comparisons. Under best conditions (including new sensors) uncertainties of 2-3%RH are achievable, but typical uncertainty is perhaps 3-5%RH for both psychrometers and capacitive sensors. Most capacitive sensors drift to higher values (except in arid conditions) so they need regular replacement or readjustment and recalibration. The observed humidities are less accurate than we would like, but they are still more accurate than short range forecasts.

In our NWP system reported humidities are converted to relative humidity, which is taken as relative to saturation over ice below 0°C (so values over 100% can arise below 0°C). Figure 3 (solid lines) shows the distribution of RH: at night there is a maximum at 100% but during the day there is a broad plateau between about 60 and 100%. There is a local minimum just below 100%, which is an artefact of reporting as dew point (in most cases) and the precision used. There is much less diurnal variation in specific humidity (not shown), but it varies by orders of magnitude with temperature and altitude.

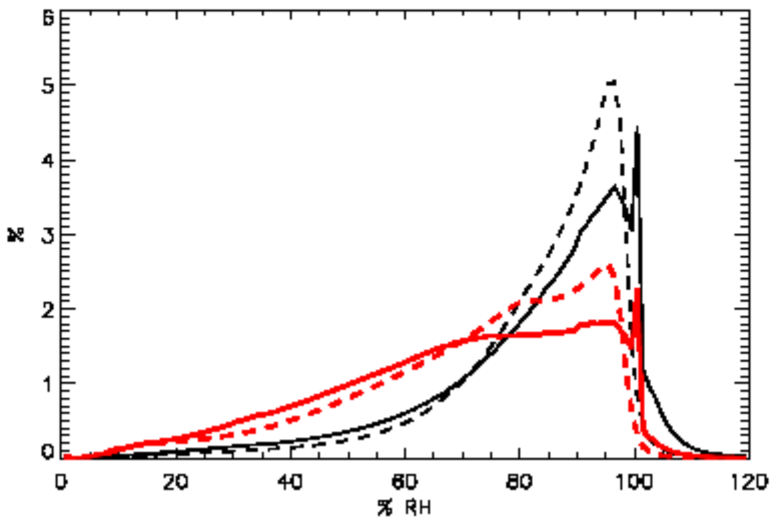


Figure 3. Distribution of RH from reports (solid lines) and from model background (dashed lines, see section 3) for Synop stations from April 2011 to March 2013. Thick red (thin black) lines are for daytime (night-time) data. See text for details.

2.2.4. Wind

Wind should be measured at 10 m height at land stations (WMO, 2010). In practice not all anemometers are at 10 m – they are more likely to be low than high (the Antarctic AWSs described by Lazzara et al, 2012, measure wind at 3 m but this is exceptional) but the resulting speed error will be less than 10% in most cases. Nearby obstacles tend to reduce the wind speed, approximate guidelines exist for raising the anemometer to compensate for this but how widely this is done is not clear. Many stations use cup anemometers to measure the wind speed. Due to cup inertia they have a non-linear response and in turbulence they tend to overestimate the mean wind speed by 1-10% (see Kristensen, 1995 and references). Sonic anemometers are used in research and have been deployed on the UK moored buoys (Turton and Pethica, 2010, and Turton, pers. comm., 2013) but probably not to many land stations. In Synop code wind speed can be reported in whole knots or m s^{-1} (48% and 52% of reports respectively; there is a factor of two error if the units indicator is wrong, which has happened but is difficult to detect in practice). Figure 4 (solid lines) shows the distribution of wind speeds reported: at night average speed is lower and the proportion of calm winds is higher.

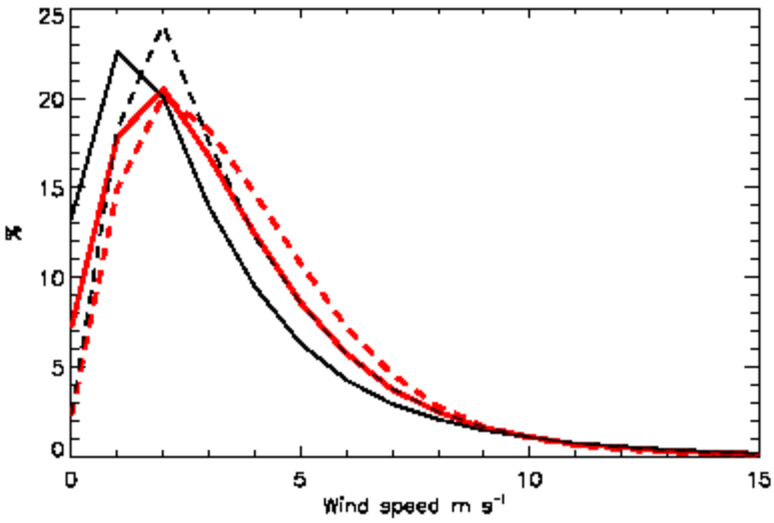


Figure 4. As figure 3 but for wind speeds (m s^{-1}) – restricted to Synop stations reporting in m s^{-1}. About 0.6% of reports are for speeds over 15 m s^{-1}.

The numbers of calm winds may be inflated by various errors (possibly including reporting of zero instead of missing data). Friction and magnetic drag can have a significant effect on the measurement of low wind speeds. Older cup anemometers used by the Met Office had start-up/stall speeds of 2.5-4.0 m s^{-1}, but starting in 1998 they were progressively replaced by anemometers that start and stall at 0.5 m s^{-1} (Sloan and Clark, 2012). Errors in anemometer calibration could cause speed biases of either sign. If an anemometer is not well maintained the friction may increase giving winds that are too weak. Makkonen et al (2012) discuss the difficulties that icing causes for wind measurement and also calibration problems.

3. Met Office assimilation system and diagnostics

3.1. Met Office global assimilation system and use of surface data in other systems

The Met Office global NWP system uses 4-dimensional variational assimilation, 4D-Var, (Rawlins et al 2007) now augmented with ensemble perturbations – discussed below. Global use of extra surface data was facilitated because the Met Office has extensive experience of using near surface temperature, humidity and wind in models for the UK area (Macpherson et al, 1996). (The UK system also uses cloud and visibility information from surface reports: Macpherson et al, 1996 and Clark et al, 2008.) The Met Office also has a global "surface" analysis which uses screen level temperature and humidity data, plus some satellite data, to update the model soil moisture field (Dharssi et al, 2011). The April 2008 change brought the usage of surface data in the global and UK analyses much closer together and some aspects of observation usage were changed in the UK and surface analyses as a result of the global experience.

Milton and Earnshaw (2007) compared global model forecasts in 2002/03 with 16 enhanced surface stations and Edwards et al (2011) compared forecasts in 2006/07 with a site in the UK. These papers describe aspects of the Unified Model (MetUM) that particularly affect the near-surface simulations. The boundary layer mixing is largely as described by Lock et al (2000), including the improvements of Brown et al (2008). The lowest wind level in the Met Office global model is at 10 m with the lowest temperature/humidity level at 20 m, skin temperature Tskin is also a model variable and is important in modelling surface fluxes. Based on the similarity theory used in the model surface exchanges temperature and humidity at a nominal screen level (1.5 m) are calculated (Edwards, 2009, 2012, describes a modification to the screen level temperature diagnostic during the evening transition with relatively weak winds). In the Met Office boundary layer formulation the surface roughness length is enhanced in hilly terrain order to represent the drag from subgrid orography (Wood and Mason, 1993). This tends to give a slow bias to the near-surface model level winds, although some attempt is made to undo this effect by diagnosing the 10 m winds without this enhanced roughness for comparison with observations (Lock and Edwards, 2012).

In 2008 the global model had mid-latitude grid spacing of about 40 km and 50 vertical levels. In November 2009 it was upgraded to 70 levels (keeping near surface levels the same) and in March 2010 the grid spacing improved to about 25 km. A six hour analysis cycle is used with the observation window centred on the nominal analysis time. Background values used in the assimilation and the monitoring statistics use bilinear horizontal interpolation to observation locations together with linear time interpolation from forecast fields output every three hours. Observation minus background (*o-b*) statistics contain contributions from both background errors and observation errors (with standard deviations σ_b and σ_o). Based on monthly *o-b* monitoring statistics some stations/variables are rejected (blacklisted) and pressure bias corrections are updated. The rejection of persistently poor stations is quite important. Real time quality control (background and buddy checks) is also performed based on the method of Lorenc and Hammon (1988) – the proportion of data rejected by the real time checks is typically O(1%) or less.

The Canadian global NWP system uses wind (only over water), pressure, temperature and humidity data from surface stations – one per six hour window (Laroche and Sarrazin, 2010, table 1). ECMWF (European Centre for Medium range Weather Forecasts) assimilates surface pressure and daytime humidity (Andersson et al, 2007) but not other surface variables from land stations; reports every 30 minutes are used if available. (For some years ECMWF assimilated island winds but this was discontinued when Kållberg, (1998, section 4) found that they were significantly weaker than the background winds and causing problems in the analysis. The forecast surface winds were probably unrealistically weak when the island winds were first assimilated.) The NCEP (National Centers for Environmental Prediction, U.S.) global system only uses pressure data from land stations (B. Ballish, 2013, pers. comm.). Various centres have a separate analysis of screen temperature and humidity, which updates soil moisture, e.g. Simmons et al (2010). Many regional NWP systems (but not all) assimilate screen temperature and Pu et al (2013) discuss some of the difficulties involved. Benjamin et al (2010) show the impact of various observation types in a regional system and discuss the use of “pseudoresiduals” to spread the effect of surface observations through the model boundary layer. Using the UK area NWP system Dow and Macpherson (2013) found that surface observations delivered the greatest benefit to the “UK Index” (mainly based on forecasts of surface variables).

The Met Office analysis system and its climatological background error covariances (Lorenc et al, 2000, Ingleby, 2001) use the terrain following model grid. For isolated surface observations 3D-Var increments are spread isotropically along near-surface levels regardless of the local orography or surface type – this is not ideal but is difficult to change directly (vertical spreading varies somewhat with latitude). Local 2D analyses of air temperature (eg Tyndall et al, 2010), are produced at several centres and can add extra detail compared to 3D analyses: background error correlations can be made a function of height difference as well as horizontal distance. 4D-Var represents anisotropic correlations to a limited extent and in July 2011 the Met Office started using a hybrid ensemble/4D-Var global assimilation system (Clayton et al, 2012) but details of the surface correlations have not been studied. Until recently the global ensemble was underspread near the surface but, partly to improve the hybrid assimilation, Flowerdew and Bowler (2013) have added vertically localised inflation and Tennant and Beare (2013) introduced perturbations to soil moisture which have improved the spread.

3.2. Representativity errors

Any physical scales, features or processes which affect the observation but are not represented in the forecast model cause representativity – or representativeness – error (Janjic and Cohn, 2006; Waller et al, 2013). (Within data assimilation systems it is usually combined with measurement error to give observation error.) Higher model resolution reduces representativity error but increases background error at model resolution (it is harder to forecast for a smaller area), typically there is a small net decrease in |*o-b*|. Because of the variation in surface properties, often at small scales, representativity is a particular issue for surface land observations (and marine observations near the coast). Some islands and peninsulas are too small to be represented in the forecast model and a direct comparison of forecast and observed wind speeds shows large biases because wind speeds are much higher over sea. Differences in surface height and surface type (agricultural, urban, snow/ice etc) can also be significant. If the observation is in a valley within an upland area then at times we may be comparing a snow-free observation with a higher-level snow-covered forecast grid point. Meteorological measurements are made over short vegetation (usually grass) where possible (WMO, 2010) – if the surrounding region is forested (or irrigated) then the observation is not typical of the region. Even in relatively “gentle” countryside there are differences between relatively close measurements (Horlacher et al, 2012). Aspects of the micro-environment (such as paved surfaces nearby) can influence measurements (Mahmood et al, 2006) – however Dow (2013, pers. comm.) found similar T and RH *o-b* statistics for UK Synop (at 1.25 m) and roadside stations (at ~2 m). Kumamoto et al (2013) found that a road affected nearby temperatures more at 0.5 m than at 1.5 m. Given all these factors it is sometimes surprising how good the agreement is between observations and model.

3.3. Model equivalents of observed variables and height adjustment

Model and observed variables need to be converted so that they are as comparable as possible. Reported wind speed and direction are converted to u- and v-components and humidity is converted to RH as discussed above. The ‘observation operator’ to calculate the model equivalent of the observed variables includes a) model diagnostics of screen variables, b) adjustments for differences between station height and model height (see below) and c) horizontal/time interpolation. Figure 5 shows a schematic with an observation at *s*, the model screen level diagnostic is at *m* calculated from values on the model levels (dashed).

Linearised versions (and their adjoints) of these observation operators are needed for the 3D/4D-Var inner loop – potentially the most complex aspect is the conversion between the main model variables and the values at the model screen level *m*. The Met Office assimilation currently uses the simple approach of applying the screen temperature (humidity) increment directly to the lowest temperature (humidity) level, Tskin is also incremented by the same amount so this corresponds to keeping the temperature difference constant.[1] This works fairly well in neutral conditions, but will give too large an increment to the lowest temperature level in stable boundary layer conditions. (In the ECMWF system Cardinali et al (1994) tried step by step linearisation of the screen temperature diagnostic but this made the operators much more complex and was not used operationally.) A linearised observation operator of intermediate complexity has been developed and is being tested. This improves consistency between the assimilation system and the forecast model diagnostic and reduces the jump in the observation penalty from one outer iteration to the next (T Payne, pers. comm. 2013). Operationally we currently only use a single outer loop – some non-linearities are treated within the inner loop (Rawlins et al, 2007; Ingleby et al, 2012).

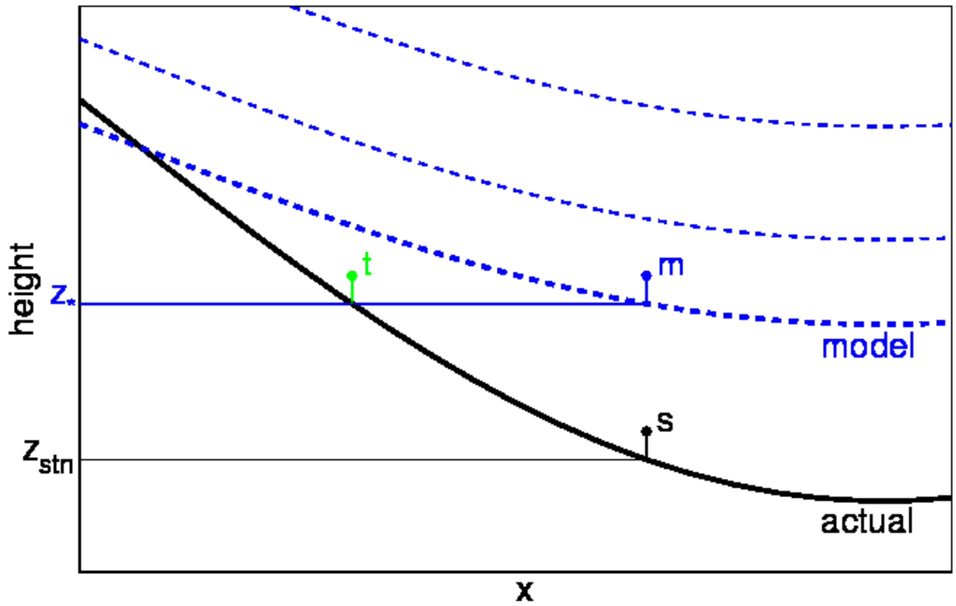


Figure 5. Schematic of height adjustment – cross-section through part of a mountainside. The thick solid line represents the real ground surface and "s" the location of the station temperature. The thick dashed line represents the model orography and "m" is the location of the model 'screen level' temperature – calculated from values at the model levels (thin dashed lines) and surface. When comparing the temperatures the difference in heights Zstn-Z* and the (estimated) lapse rate has to be taken into account. "t" is a hypothetical measured temperature at height Z*.

For convenience reported values are adjusted to the model height – for linear adjustments this is equivalent to adjusting the model values to the station height. The Met Office uses a lapse rate correction of -0.0065 °C m^{-1} (as used in the International Standard Atmosphere). Height adjustments for humidity and wind speed were introduced in 2008. Figure 6 shows *o-b* statistics for the northern extratropics as a function of Δz – the values shown are after height adjustment except that wind speed is shown both before and after. Individual station statistics (not shown) display a lot of scatter in such plots, but when averaged there are some clear patterns. For temperature the height adjustment is essential for larger Δz values although on average the

[1] The temperature increment calculated for the bottom atmospheric level is also applied to the top soil level (although perhaps it should be scaled rather than applied in full). The increment is applied to the soil temperature even if snow is present, so long as it doesn't raise the temperature above freezing (B Macpherson, pers. comm. 2013).

correction is slightly too large and -0.0055 °C m^{-1} might be better. There is some seasonal and diurnal variation with lower lapse rates applicable in winter and particularly at night (tables in next section) reflecting greater stability then. Temperature lapse rates also affect the surface pressure differences (Figure 6a). For both variables the standard deviations (SDs) tend to be highest in winter and lowest in summer.

For RH an adjustment of 0.001% m^{-1} was introduced in 2008 and on average this gives reasonably unbiased results (Figure 6c).[2] However the background values are slightly wetter than the observations (especially in spring, see section 4.2) – probably because precipitation in the forecast is slightly high (Milton and Earnshaw, 2007; Ingleby et al 2012). Specific humidity (not shown) has a slope of the opposite sign to RH. At night the model distribution of RH (Figure 3) is more peaked than the observed distribution and the peak is at about 95% rather than 100%; during the day the model distribution is much more peaked than observed.

For wind speed and direction statistics there is a marked difference between stations below and above the model orography with both showing a change in bias at about Δz=200 m. For stations above this the reported winds are increasingly strong and veered compared to the model winds. There is less scatter for stations more than ~200 m below model orography - presumably these are valley stations with weaker winds on average - there is some asymmetry because wind speed cannot drop below 0 m s^{-1}. The wind speeds are adjusted using $s' = s / S$ where S=1 for $\Delta z < 100$, $S=1+0.002\Delta z$ for $100 < \Delta z < 1100$ and S=3 for $\Delta z > 1100$. No adjustment is applied to wind direction. This form of correction is based on Howard and Clark (2003, 2007) and the coefficients from global model statistics for 2006 and 2007. It was decided to use a generic correction (based on Δz only) rather than a station by station correction. The adjusted wind speeds show much less bias. The wind speed difference SDs tend to be largest in winter, when the speeds are largest.

3.4. Variations in lapse rate

It is clear that the temperature lapse rate varies with season (Figure 6 and Rolland, 2003) and also to some extent with the diurnal cycle and synoptic conditions. Post-processing UK area forecasts Sheridan et al (2010) used a template of model grid points near each location to calculate a lapse rate by regression of 20 m temperature against surface height, they use this for $|\Delta z|$ up to 70 m (earlier work using the model vertical profile above each location found an excessive sensitivity to the levels used in the computation). There is an assumption that cloud conditions etc are similar within the local area used. Results using such a lapse rate in the global observation processing suggest that whilst it is easy to improve the temperature *o-b* fit for small $|\Delta z|$ in stable conditions it is much more difficult to provide a global solution that improves the fit in these conditions whilst not degrading it in neutral and unstable conditions or for large $|\Delta z|$. A possible alternative would be to use a lapse rate that is a function of latitude, month and day/night.

McCutchan (1983) and McCutchan and Fox (1986) compared temperature, humidity and wind from mountain stations and nearby radiosondes. Pepin and Seidel (2005) compared temperatures from high-level stations with those from the NCEP/NCAR reanalysis. The general conclusions are that the temperature lapse rate can be different

[2] Preliminary work looking at similar statistics from the ECMWF suggests that the ECMWF model has a different behaviour of RH biases as a function of Δz (and a dry bias overall) so that this adjustment should not be taken as universally applicable. In contrast the lapse rate adjustment for temperature should be generally applicable.

following a mountain slope to that in the free air nearby. The aspect of the slope, time of day (especially night-time inversions) and local wind strength can all have an effect on the comparisons. Pepin and Seidel (2005) found very high anomaly correlations between mountain-top and free-air temperatures and "most of the low correlations are from deeply incised mountain valley locations". The mountain slope lapse rate should be more appropriate than the free air lapse rate when adjusting from the reported screen T to model height (in Figure 5 the temperature at **t** should be a better match to the model screen temperature than the free air temperature at **m**). The characteristics of cold pools (that form in valleys at night under weak wind conditions) are discussed by Sheridan et al (2013).

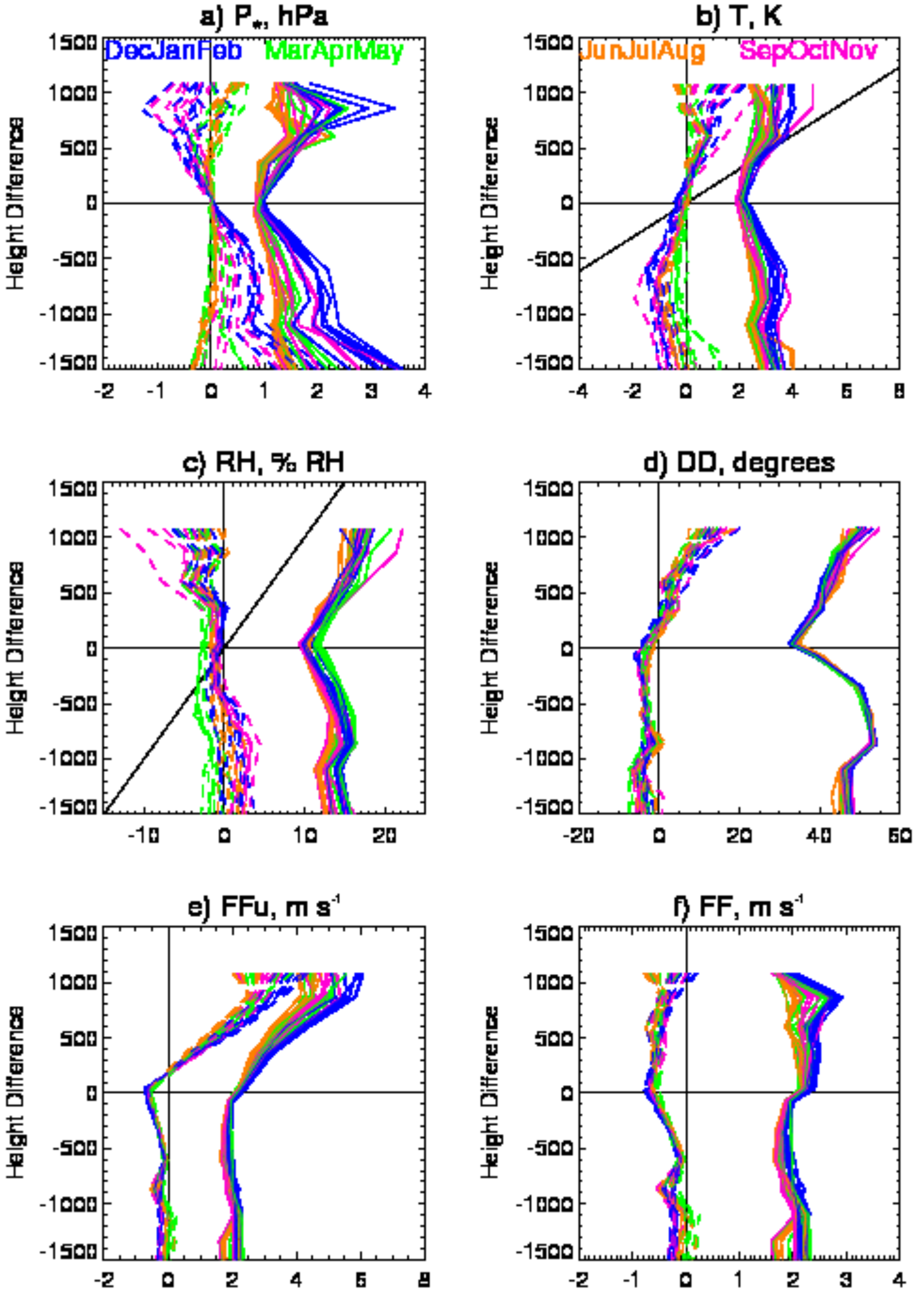


Figure 6. Synops for 20°-90°N, *o-b* statistics as a function of Δz (in 250 m bins): April 2011 – March 2013, individual months plotted, colour coded as indicated at top. Dashed

lines give biases, solid lines standard deviations. a) surface pressure, b) temperature, c) relative humidity, d) wind direction, e) unadjusted wind speed, f) adjusted wind speed. For temperature and RH the diagonal black lines show the height adjustment applied.

3.5. Summary o-b statistics

Tables 3 to 7 provide *o-b* statistics for pressure, temperature, RH, wind speed and direction respectively – all after application of height adjustments. The statistics have been computed station by station using threshold checks on |*o-b*| (15 hPa for pressure – possibly over-generous, 15°C for temperature, 50% for RH, 15 m s^{-1} for wind speed and 100° for wind direction) and then aggregated. The statistics include some stations/variables that are not assimilated operationally. We try to detect stations with large σ_o/σ_b ratios and omit them from the assimilation (either as a class or as individual stations). The main tool for doing this is looking at SD(*o-b*), which is $\sqrt{(\sigma_o^2 + \sigma_b^2)}$. In some cases we can assume that σ_b is approximately constant but for others more judgement is required.

For pressure (Table 3) there is little difference between the day and night statistics except for those stations well above or below model level (which have larger errors). For the bulk of stations (with $|\Delta z|<250$m) the errors are quite small with negligible biases and SD o-b of about 0.5 and 0.4 hPa for Synop and Metar respectively. (The assumed observation errors of 1 hPa are overestimates, see section 5.4.) Metar data were not being assimilated for the first few months of these statistics, this might make the Metar statistics slightly worse. On the other hand Metars tend to be in densely populated areas, whereas some Synops are in remote areas which may have larger σ_b. Metar pressures may be subject to feedback from aircrew which will help to reduce errors.

	Day				Night			
	N'000	**Mn o**	**Mn o-b**	**SD o-b**	**N'000**	**Mn o**	**Mn o-b**	**SD o-b**
Synop								
Ocean	2267	1012.05	-0.01	0.77	2149	1012.07	0.02	0.75
Low	2998	879.35	0.15	0.73	2698	881.31	0.48	0.91
Mid	19486	976.38	-0.00	0.48	18248	977.60	0.07	0.54
High	975	927.25	-0.08	0.65	907	927.89	-0.23	0.74
Metar								
Ocean	2693	1013.40	0.01	0.45	2132	1013.39	0.09	0.46
Low	3671	882.71	0.12	0.66	3041	882.40	0.36	0.70
Mid	44206	979.35	0.00	0.39	40726	979.89	0.05	0.40
High	340	926.59	-0.34	1.15	264	925.60	-0.54	1.01

Table 3. Pressure (hPa) statistics for April 2011 – March 2013, split by day and night. N'000 gives the number of reports in 1000s. Mean (Mn) and Standard Deviation (SD) of o-b. "Ocean" denotes stations at model sea points, other stations are categorised by Δz=Zstn-Z*: low $\Delta z<-250$m; mid $|\Delta z|<250$m and high $\Delta z>250$m. σ_o is taken to be 1.0 hPa for most stations (it is slightly increased for Metars reporting in whole hPa, see Appendix 1).

For temperature (Table 4) and RH (Table 5) the Synop and Metar statistics are broadly similar, *o-b* differences are larger at night for temperature but only slightly so for RH. As for pressure the σ_o estimates are too large. The *o-b* SDs for stations at model sea points are fairly small but the mean differences have a diurnal cycle from about 1 to -1°C and about -5 to 2%RH. For $|\Delta z|<250$m o-b temperature SDs are generally between 0.5 and 0.8 °C between 60°S and 60°N but they are larger at high latitudes, being up to 2 °C or more over Antarctica. At night the biases for temperature and RH (and wind direction, below) become larger for $\Delta z > 250$ m so these values aren't assimilated, apart from this we assimilate data with $|\Delta z| < 500$ m. Biases between screen temperatures and background values are much less constant than those for pressure and have diurnal (Qin et al, 2010), seasonal and synoptic components. Forecast bias is also more significant than for pressure, so we do not attempt to bias correct temperatures apart from the height adjustment described above.

	Day				**Night**				
	N'000	**Mn o**	**Mn o-b**	**SD o-b**	**N'000**	**Mn o**	**Mn o-b**	**SD o-b**	**Nstn**
Synop									
Ocean	2393	19.58	0.91	0.71	2248	16.50	-1.05	0.99	640
Low	3252	10.90	-0.29	1.18	2913	4.09	-0.78	1.60	925
Mid	22707	14.44	0.18	0.78	21161	7.50	-0.63	1.31	6929
High	1115	13.68	0.18	1.53	1030	8.89	1.14	1.64	259
Metar									
Ocean	2679	21.82	1.11	0.76	2118	18.87	-1.13	0.98	250
Low	3660	12.69	-0.24	1.25	3039	5.59	-0.96	1.49	371
Mid	44092	16.84	0.24	0.73	40478	10.38	-0.73	1.10	3648
High	306	17.60	0.27	2.13	214	13.02	1.10	2.18	40

Table 4. As table 3 but for temperature (°C). σ_o is taken as 2.0 (2.1) °C for Synop (Metar). The right hand column gives the number of stations in each category.

	Day				**Night**			
	N'000	**Mn o**	**Mn o-b**	**SD o-b**	**N'000**	**Mn o**	**Mn o-b**	**SD o-b**
Synop								
Ocean	2244	75.83	-4.20	4.65	2096	82.34	2.78	4.65
Low	3221	67.31	-1.63	6.17	2885	81.72	-0.90	6.49
Mid	22030	66.61	-2.23	4.82	20464	80.62	-1.07	5.04
High	1056	67.33	-0.74	6.09	950	76.57	-5.54	6.91
Metar								
Ocean	2660	72.34	-6.45	4.71	2108	79.26	1.50	4.82
Low	3640	62.27	-1.64	5.54	3023	77.52	-1.10	6.67
Mid	44003	63.12	-3.08	4.41	40405	77.40	-1.66	5.24

	Day				Night			
	N'000	Mn o	Mn o-b	SD o-b	N'000	Mn o	Mn o-b	SD o-b
High	299	58.83	-1.76	7.48	201	70.27	-4.87	7.10

Table 5. As table 3 but for relative humidity (%). σ_o is taken as 10.0 (11.0) % for Synop (Metar).

For wind speeds (Table 6) the reported speeds are much stronger at the "ocean" stations (model sea points) but the background there is even stronger giving significant biases. At "land" stations the winds are weaker at night and the background winds are 0.5 to 0.7 m s^{-1} stronger than the winds for stations with $|\Delta z|<250$m. The Metar winds are slightly stronger than Synop winds on average and in slightly better mean/SD agreement with the background. For wind direction (Table 7) there is a modest negative bias at most stations (slightly larger at night and in winter). The SDs are about 10° for "ocean" stations and those with $|\Delta z|<250$m, but significantly larger for stations with large Δz; there is a local maximum in the tropics (where the speeds have a minimum) and very large values at high southern latitudes.

	Day				Night			
	N'000	Mn o	Mn o-b	SD o-b	N'000	Mn o	Mn o-b	SD o-b
Synop								
Ocean	2384	5.10	-1.24	1.55	2245	4.82	-1.87	1.79
Low	3176	2.32	-0.22	0.98	2843	1.71	-0.25	0.91
Mid	21307	3.37	-0.35	1.06	19781	2.70	-0.74	1.09
High	1085	2.46	-0.60	1.13	1000	2.60	-0.04	1.36
Metar								
Ocean	2683	5.03	-1.17	1.12	2125	4.38	-2.19	1.46
Low	3692	2.87	0.32	0.95	3059	2.15	0.13	0.93
Mid	44352	3.80	0.12	0.79	40840	2.85	-0.52	0.91
High	354	2.65	-0.76	1.23	266	2.65	-0.35	1.51

Table 6. As table 3 but for wind speed (m s^{-1}). σ_o is set as 2.0 m s^{-1} for both u- and v-components of Synop and Metar data.

	Day			Night		
	N'000	Mn o-b	SD o-b	N'000	Mn o-b	SD o-b
Synop						
Ocean	2232	-1.43	10.15	2078	-2.42	12.25
Low	2339	-3.57	14.15	1940	-2.91	16.20
Mid	19038	-3.29	10.27	16979	-4.74	12.75
High	891	3.29	13.07	799	6.67	16.44
Metar						

	Day			Night		
	N'000	**Mn o-b**	**SD o-b**	**N'000**	**Mn o-b**	**SD o-b**
Ocean	2433	-0.56	9.52	1893	-3.37	11.61
Low	2533	-3.46	15.45	1994	-4.15	17.44
Mid	38594	-3.80	8.57	33553	-6.22	10.50
High	283	5.85	15.05	212	7.14	22.82

Table 7. Similar to table 3 but for wind direction (°). The numbers differ from those in table 6 because the |*o-b*| checks are performed separately for speed and direction.

	Manned				Automated			
	N'000	**Mn o**	**Mn o-b**	**SD o-b**	**N'000**	**Mn o**	**Mn o-b**	**SD o-b**
Synop								
P (hPa)	18302	975.08	0.05	0.51	13864	978.86	-0.02	0.44
T (C)	18791	13.56	-0.10	0.88	19380	7.39	-0.36	0.90
RH (%)	18677	71.05	-1.42	4.49	18132	76.40	-1.67	3.87
FF (m/s)	18665	2.60	-0.71	0.98	16651	3.84	-0.23	1.14
DD (°)	15772		-3.89	11.25	15135		-3.61	10.41
Metar								
P (hPa)	30248	981.25	0.03	0.42	39957	974.62	0.03	0.37
T (C)	30387	16.03	0.02	0.77	39541	11.93	-0.47	0.71
RH (%)	30313	72.21	-2.88	4.52	39487	69.49	-1.45	3.93
FF (m/s)	30386	3.41	0.02	0.80	40041	3.30	-0.43	0.85
DD (°)	25527		-4.02	10.03	34140		-5.82	7.33

Table 8. Similar to table 3 but comparing manned and automated stations (at model land points, with |Δz|<250m) for different variables: day/night combined. For Synop/Metar temperature there were 4238/1818 manned stations and 2162/1186 automated stations (stations with at least 90% manned/automated reports respectively) and 528/641 mixed stations (latter not shown).

Table 8 compares statistics for manned and automated stations, from the mean temperature the automated stations are clearly at higher latitudes on average. There are many automated Synops in Europe (especially northern Europe) and Canada, there are moderate numbers in the Pacific (including Australia and New Zealand), USA, Japan and South Africa and some in Antarctica. Automated Metar distribution is similar but with very large numbers in the USA. Reports from some newer AWSs are only available in BUFR format (not Synop) and are not currently processed in our system. Pressure and RH SD (*o-b*) are somewhat better for automated stations; this is partly because of the avoidance of typographical errors, other factors may also affect RH. Temperature differences tend to be larger at high latitudes as noted above. The wind speed SD (*o-b*) is somewhat larger for automated stations, but so is the mean wind; the bias is worse for manned stations (reflecting problems in parts of the tropics).

For the same two year period statistics for UK and Eire (block “03” Synop stations) have been compared between the global model and the UKV model which has a grid spacing of 1.5 km over the British Isles. The most marked results are that temperature SD (*o-b*) is lower for the UKV: 1.04 vs 1.28 °C and that mean background wind speeds are lower for the UKV: 4.87 vs 5.46 m s^{-1}. The latter result is slightly surprising as higher resolution might be expected to give slightly higher wind speeds, but the numerical diffusion applied is necessarily different and the UKV has a more sophisticated surface scheme (with different ‘tiles’ for different vegetation types within each model grid box, in the global model only the dominant type is used in each grid box).

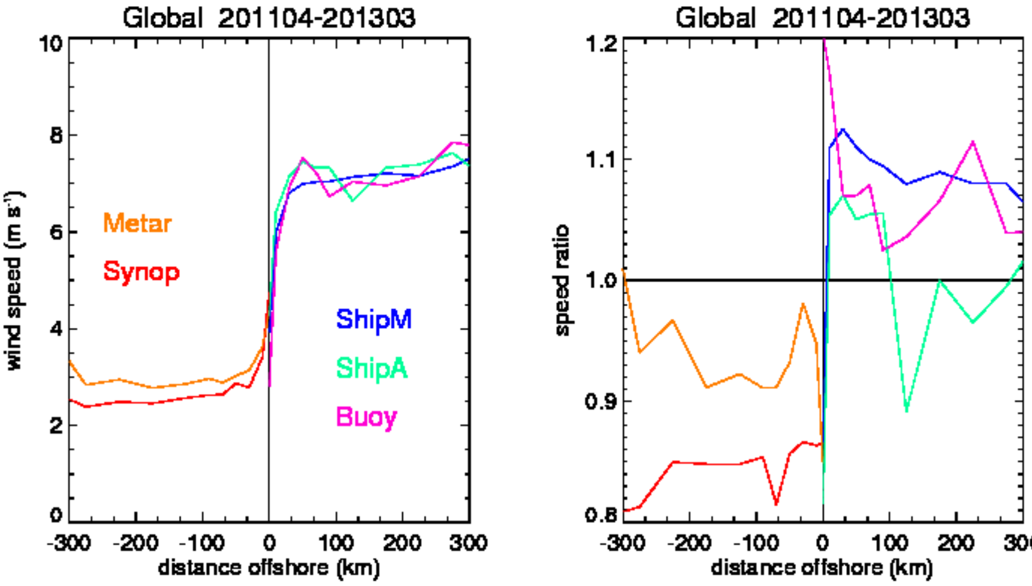


Figure 7. Left: mean wind speed as a function of distance offshore for April 2011 to March 2013 for Synops, Metars, Manual/Automated Ship reports and moored Buoys reporting in Ship code – marine speeds adjusted to 10 m as appropriate. (A ~10 km land-sea mask was used, negative values are over land; statistics in 20 km bins within 100 km of the coast, and then 50 km bins out to 300 km, reports further from the coast plotted at ±300 km and reports over the “wrong” surface plotted at 0 km). Right: (mean observed speed)/(mean background speed) – figure 4 of Ingleby (2010) showed earlier speed ratios for marine data only.

Over land mean 10 m wind speeds are 2.5 - 3 m s^{-1}, over the ocean they are 7 – 7.5 m s^{-1} as shown in figure 7 (these figures relate to observation locations and hence are biased towards northern mid-latitudes) with most of the adjustment being within a few tens of kilometres of the coast.

Figure 8 shows that background daytime wind speeds over North America are quite realistic, although they are not strong enough over the eastern slopes of the Rockies and they are slightly too strong over the eastern USA. At night the background winds over the western USA/Rockies are generally reasonable but over the eastern USA the background winds are clearly too strong. This seems to be a specific example of a feature in Table 6: negative nighttime speed biases for stations with $|\Delta z|<250$m and smaller biases for stations in more mountainous regions. The model has excessive boundary layer mixing at night which probably makes the wind speeds too strong - Brown et al (2008) were able to reduce this over the ocean, but over the land the mixing is currently necessary in order to avoid excessively low temperatures in valleys under certain conditions. In parts of the tropics there are biases in daytime wind speeds as well, particularly over southern Asia (Figure 9) although other areas are affected as well. The daytime differences may be a combination of model and observation errors. There is some seasonal variation in the speed biases in this region and a slight increase in the biases in recent years (not shown).

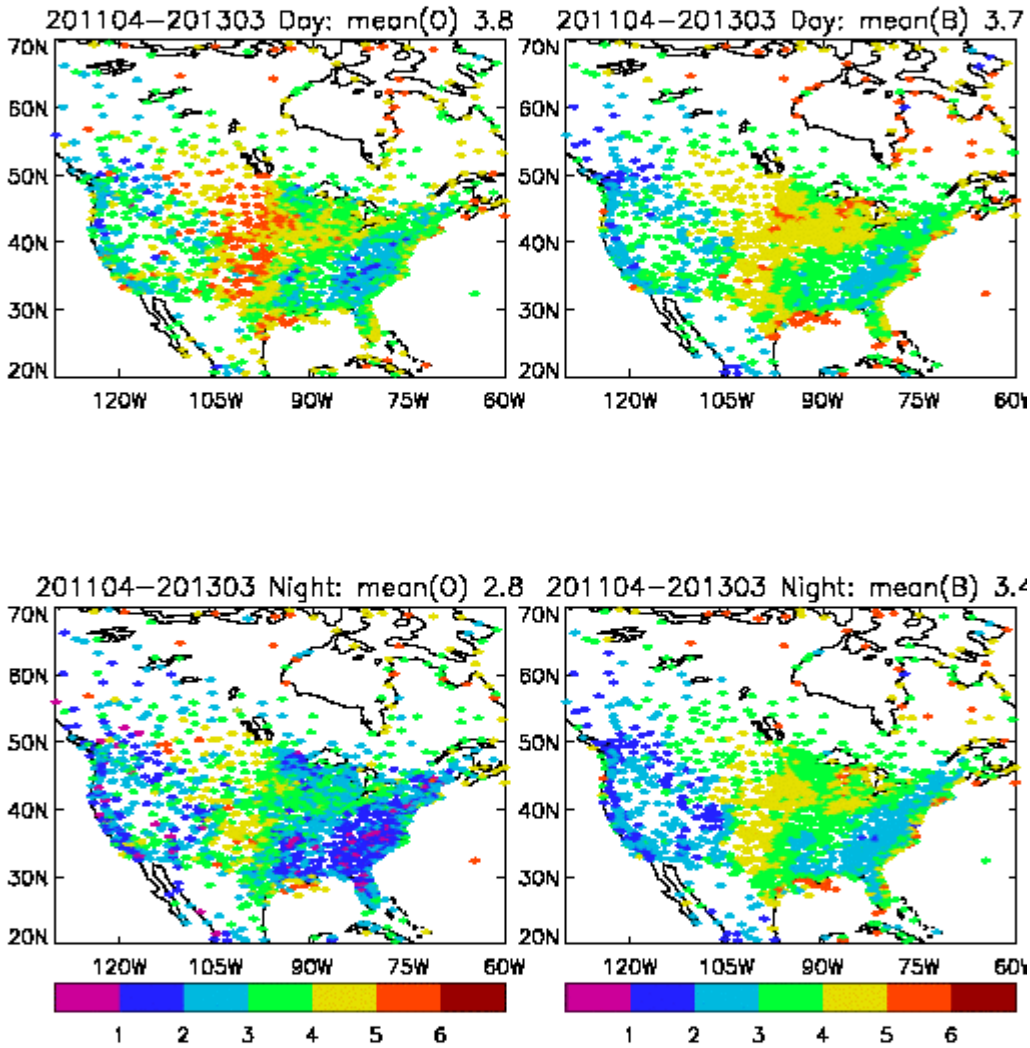


Figure 8. Wind speeds (m s^{-1}) at Metar stations for the observations (O) and background (B), day and night separate, mean over April 2011 - March 2013. The numbers in the plot titles give the mean over the stations shown.

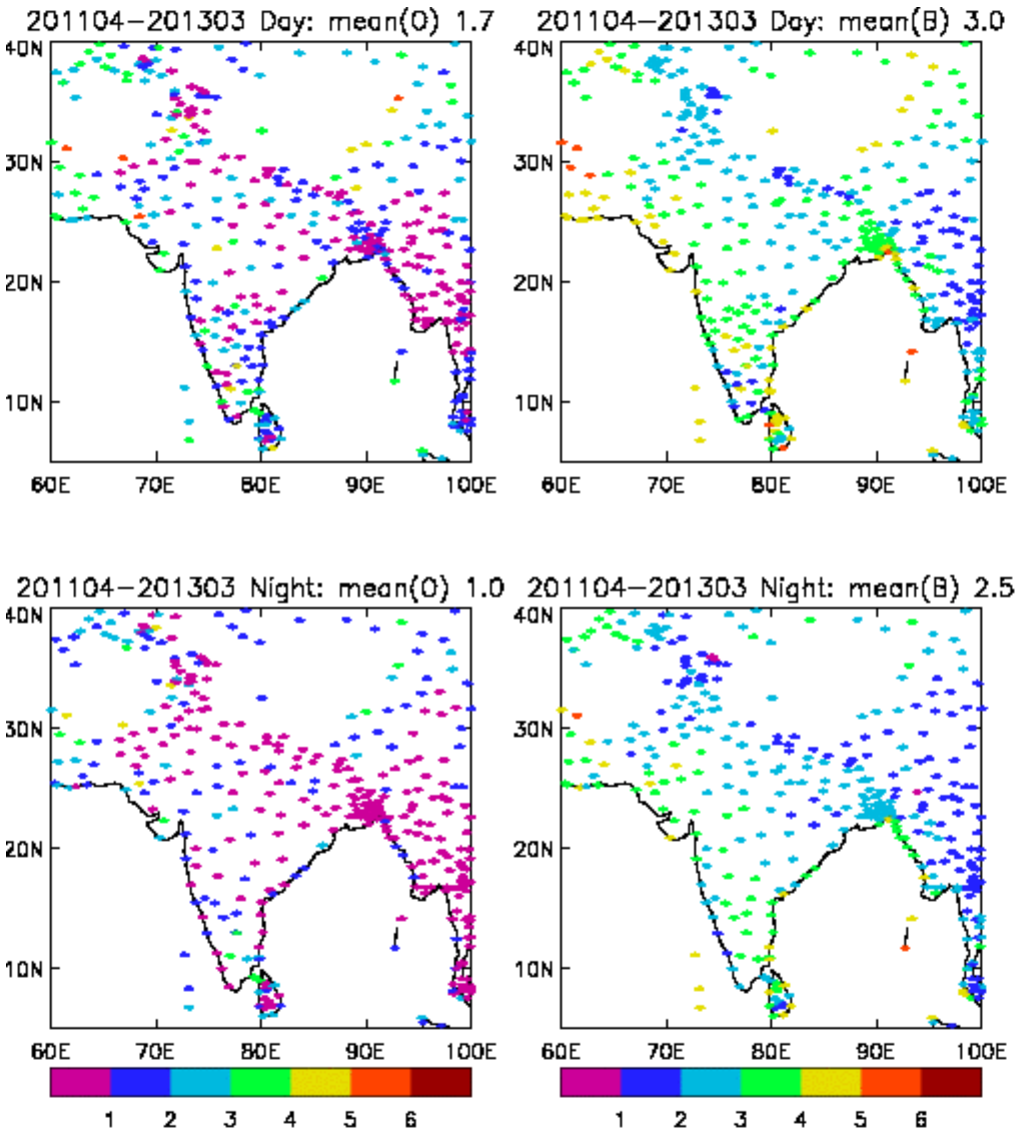


Figure 9. As figure 8 but for Synops over southern Asia.

3.6. April 2008 changes

Prior to April 2008 surface pressure reports were assimilated globally plus marine wind reports (surface pressure reports were/are used regardless of |Δz|, but they can be rejected or bias corrected based on monitoring statistics). Following this change most T, RH and wind reports from surface stations (land and marine - but not for drifting buoys) were assimilated. The Synop T, RH and wind data are not used if |Δz| > 500 m, at night this is modified slightly so that data with Δz> 250m are also rejected (see above, at night the boundary layer becomes shallower and stations on hills may be sticking out above the main boundary layer). Synop winds are not assimilated at model sea points or between 30°S and 30°N because of the biases in O-B wind speeds (the 30°S-30°N rejection is somewhat arbitrary and could be refined).

The changes above were tested and refined in a series of 3D-Var trials (with ~60/~120 km forecast/analysis grid spacing). The assimilation of Synop T and RH gave an immediate improvement. On first attempt the assimilation of Synop winds gave a degradation; after the removal of island and tropical winds it gave a slight improvement.

In contrast T and RH from unresolved islands gave a slight beneficial impact in forecast trials. Early trials rejected T, RH and wind with |Δz| > 200 m, but broadening the Δz range gave a slight improvement. For some stations the proportion of calm reports seems suspiciously large, but removing all calm Synop winds gave slightly worse results (a stuck value check was also tested, but its main effect was to reject calm winds). The inclusion of ship and buoy T and RH gave fairly neutral results. It was found that the (strong katabatic) winds from some Antarctic stations were giving systematic pressure dipole increments and it seemed likely that the wind differences were related to orographic features not resolved by the model. To remove these increments a vector difference criterion was added to the wind blacklisting criteria.

Overall forecast results for June 2006 gave a consistent positive impact, with rms (root mean square) Northern hemisphere scores improved by 0.7% on average - impact on the Southern hemisphere is smaller as expected, because of the smaller number of stations there. A December 2006 trial gave neutral impacts in the Northern Hemisphere and slightly negative impacts in the Southern Hemisphere. Screen T and RH performance was improved, particularly at short range, in both trials. See Section 4 for later results. Most 3D-Var tests were performed using just one report per Synop station per 6-hour window. Using hourly reports (where available) gave neutral or slightly negative results. (The extraction window for Synop data was extended from 3 to 6 hours.) A short 4D-Var trial showed a small positive impact from using hourly Synop results. If reports from a particular station are available every 20 or 30 minutes they are thinned to hourly before assimilation.

4. Impact of surface data and performance of forecasts since 2007

The impact of the April 2008 assimilation changes has been briefly discussed. Here we present quantitative results for the impact of surface data and also look at the evolution of surface verification and *o-b* statistics since 2007. To understand the statistics since 2007 the other main changes relevant to surface performance are summarised.

4.1. Data addition/denial results

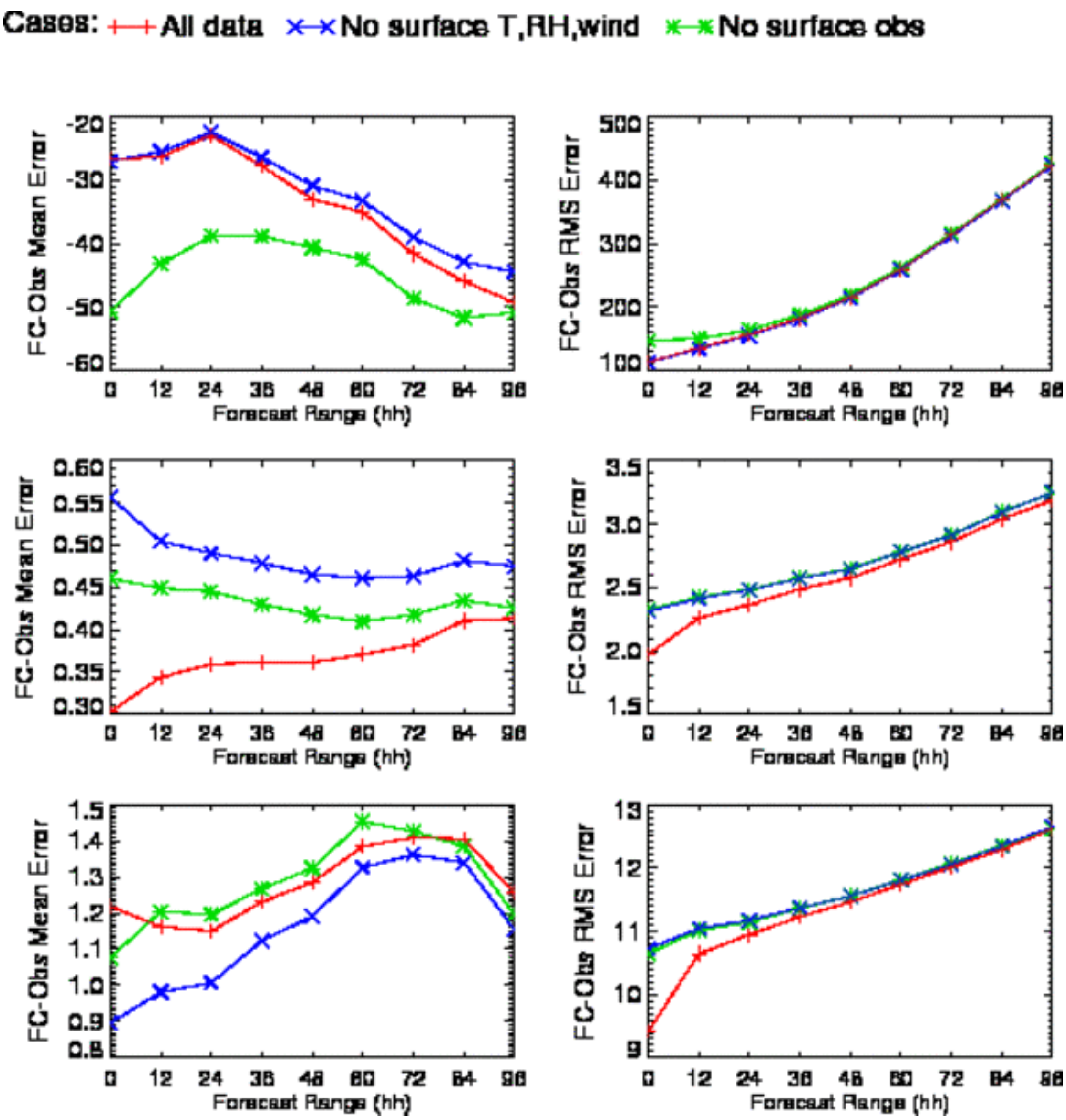


Figure 10. Verification results for 20-90°N for December 2008: mean (left) and rms (right) differences. Forecasts from both 0000 and 1200 UTC, verified every twelve hours. Top: Pmsl (Pa), middle: screen temperature (K), bottom: screen RH (%) verified against Synops, ships and buoys. Results are available to a range of 144 h, but are only shown to 96 h. See text for details of the runs.

In late 2009 4D-Var data denial experiments were run for 1-31 December 2008 and 24 June to 24 July 2009. The control was based on the operational system at the time and there were trials a) without surface temperature, humidity and wind and b) removing surface pressure as well (labelled “no surface obs”). The experiments used 70 vertical levels and a horizontal grid spacing of ~40/~120 km for the forecasts/analyses. The results were generally consistent with the earlier 3D-Var trials. The assimilation of surface T, RH and wind (which are relatively accurate observations with dense coverage over Europe in particular) causes the variational minimisation algorithm to converge slightly more slowly. Selected rms verification results are tabulated in Tables 9 and 10 and statistics for the northern extratropics in December 2008 are shown in Figure 10. In the extratropics the forecast errors are largest (especially for Pmsl) in winter and the impact of surface T, RH and wind data on surface temperature forecasts is slightly larger in winter (Tables 9 and 10). There is little impact on surface wind forecasts although some of the rms scores are fractionally reduced. For some regions/variables the assimilation of extra data gives slightly worse scores, this may be due to sampling error (related to the length of the trials), problems with observation quality or to sub-optimal correlations between different variables in the assimilation system.

In Figure 10 the biases are small relative to rms differences but for temperature the "All data" run has the lowest biases. The June/July 2009 rms results for the northern extratropics are qualitatively similar to those in Figure 10 except that for temperature the "All data" run shows no real improvement beyond T+48 unlike the small improvement in December 2008. The main benefit of assimilating surface T, RH and wind comes in improved surface T and RH forecasts up to two days range - there is a slight improvement to short range pressure forecasts but too small to be seen in the figure. Any benefit on short range T and RH forecasts from pressure assimilation is very small. The impact of surface pressure assimilation on the pressure forecasts (about 5% reduction in rms at T+24) appears smaller in Figure 10 than might have been anticipated - but other in situ and satellite data contribute to the surface pressure analysis. In the tropics and southern extratropics the impact of surface pressure data extends further into the forecast (not shown). Because of the use of in situ pressure data in the calibration (bias correction) of satellite soundings (Healy, 2013) the results here will underestimate the long-term impact of the pressure data. As measured by the "Global NWP Index" (Appendix of Rawlins et al, 2007, based on verification scores for Pmsl, 500 hPa height and 250 and 850 hPa winds) the impact of surface pressure data is three or four times that of the surface T, RH and wind. Surface pressure has more impact on upper atmosphere variables (including on height verification up to 100 hPa – not shown).

	Northern Extratropics			**Tropics (20°S-20°N)**			**Southern Extratropics**		
Range	**NoSf**	**Nothw**	**All**	**NoSf**	**Nothw**	**All**	**NoSf**	**Nothw**	**All**
Pmsl:									
T+24	*1.64*	1.55	1.56	*1.25*	1.21	1.21	*1.39*	1.31	1.30
T+48	*2.18*	2.14	2.15	*1.32*	1.27	1.27	*1.75*	1.68	1.68
T+72	*3.16*	3.13	3.14	*1.39*	1.32	1.33	*2.28*	2.21	2.19
T:									
T+24	2.49	2.48	**2.37**	1.75	1.75	**1.73**	2.21	2.20	2.20
T+48	2.65	2.65	**2.57**	1.79	1.80	1.79	2.27	2.27	2.26
Wind:									
T+24	3.10	3.10	3.09	3.01	3.01	3.01	3.38	3.36	3.36
T+48	3.33	3.33	3.32	3.10	3.10	3.10	3.61	3.57	3.57

Table 9. Selected verification results for December 2008 for experiments without surface data (NoSf), without surface T, RH and wind (Nothw) and with all data (All). Rms verification vs surface observations for Pmsl (hPa), temperature (K) and vector wind (m s^{-1}). Values 1% better (worse) than those for Nothw are in **bold** (*italics*).

	Northern Extratropics			Tropics			Southern Extratropics		
Range	**NoSf**	**Nothw**	**All**	**NoSf**	**Nothw**	**All**	**NoSf**	**Nothw**	**All**
Pmsl:									
T+24	*1.23*	1.16	1.15	*1.29*	1.24	1.25	*1.57*	1.41	1.42
T+48	1.59	1.58	1.57	*1.34*	1.32	1.32	*2.19*	2.01	2.02
T+72	2.11	2.09	2.07	*1.41*	1.39	1.39	*2.95*	2.90	2.91
T:									
T+24	2.23	2.22	**2.18**	1.85	1.84	**1.82**	2.32	2.31	**2.26**
T+48	2.39	2.39	2.37	1.88	1.87	1.86	2.39	2.38	**2.35**
Wind:									
T+24	2.96	2.95	2.94	3.07	3.08	3.08	3.24	3.23	3.22
T+48	3.20	3.19	3.18	3.17	3.17	3.16	3.41	3.40	3.39

Table 10. As Table 9 but for June/July 2009.

	N. Extratropics		Tropics		S. Extratropics	
Range	**Cntl**	**Metar**	**Cntl**	**Metar**	**Cntl**	**Metar**
Pmsl:						
T+24	1.65	1.64	1.30	**1.29**	1.45	1.44
T+48	2.26	**2.23**	1.41	**1.39**	1.80	1.81
T+72	3.22	**3.17**	1.55	**1.50**	2.29	*2.32*
T:						
T+24	2.76	2.76	1.84	1.84	2.50	2.49
T+48	3.01	3.01	1.90	1.90	2.58	2.58
Wind:						
T+24	3.26	3.25	3.05	3.05	3.51	3.50
T+48	3.52	3.50	3.15	3.15	3.70	3.70

Table 11. Similar to table 9 but for December 2010: control and trial assimilating Metar data. **Bold** (*italics*) highlight scores where the trial is at least 1% better (worse).

In early 2011 a control and trial were run for 1-31 December 2010, based on the 4D-Var system operational at the time but with a forecast/analysis grid spacing of ~60/~120 km. The trial included assimilation of Metar data (and also a minor change to the quality control of radiosonde humidity – this will have had little impact, especially at the surface). The results (Table 11) show improved Pmsl forecasts in the northern extratropics and tropics, but not in the southern extratropics. The impact on other surface variables was largely neutral (in the northern extratropics the surface temperature forecasts were slightly cooler - in better agreement with Synops – and wind speed rms is slightly better from T+72 onwards). Note that adding Metar observations will result in a slightly worse analysis fit to Synop data and that this might adversely affect the verification figures at short range (Metar data were not used in the verification).

Metars were monitored from January 2011 and assimilated in the global 4D-Var from July 2011. They were included in the surface (soil moisture) analysis from March 2012. There is a duplicate check so that if there is a collocated Synop station the Metar reports are not assimilated (this rejects approximately half of Metar stations). Apart from the duplicate check and pressure processing Metar usage is the same as for Synops.

4.2. Operational near-surface performance since 2007

In April 2008 as well as the assimilation changes mentioned above improved soil properties were introduced (Dharssi et al, 2008) – the soil change improved screen temperature and RH forecasts at all ranges. In November 2008 a snow analysis was introduced (Pullen et al, 2010), minor changes were made to marine surface pressure observation errors, and also various changes to the boundary layer including the screen temperature diagnostic (Edwards, 2012).

In March 2010 the operational global model grid spacing was improved to ~25km and there were changes to the use of marine surface data based on Ingleby (2010) including tighter quality control, better adjustment of marine winds to 10 m and assimilation of air temperature data from Arctic buoys. In July 2010 assimilation of ASCAT surface soil wetness (Dharssi et al, 2011) was implemented. At various times extra surface data were introduced by improvements to metadata, use of “mobile” Synops (March 2009) and processing extra humidity data (either reported as RH or reported without a pressure), but the largest tranche of “new data” came from Metars in July 2011.

Figure 11 shows operational verification against surface temperature observations for the northern extratropics and tropics. An improvement to the Met Office scores can be seen in April 2008 together with slight improvements since. Up to T+24 the improvement comes both from the assimilation of more surface data and from the changes to soil properties, at longer ranges the soil changes dominate. In the extratropics there is a large seasonal cycle to the rms difference. Despite the generally very good T+24 forecasts December 2012 results were rather poor – this is thought to be due to a warm bias over snow when there is large scale cold advection. For surface relative humidity and wind (not shown) there is also an improvement in April 2008 and the Met Office analyses/forecasts fit the observations better than alternative NWP systems. For wind the improvement is clearer for the tropics – but cannot be due to the assimilation of surface winds there. For pressure (not shown) ECMWF performs best in the extratropics (Met Office in the tropics).

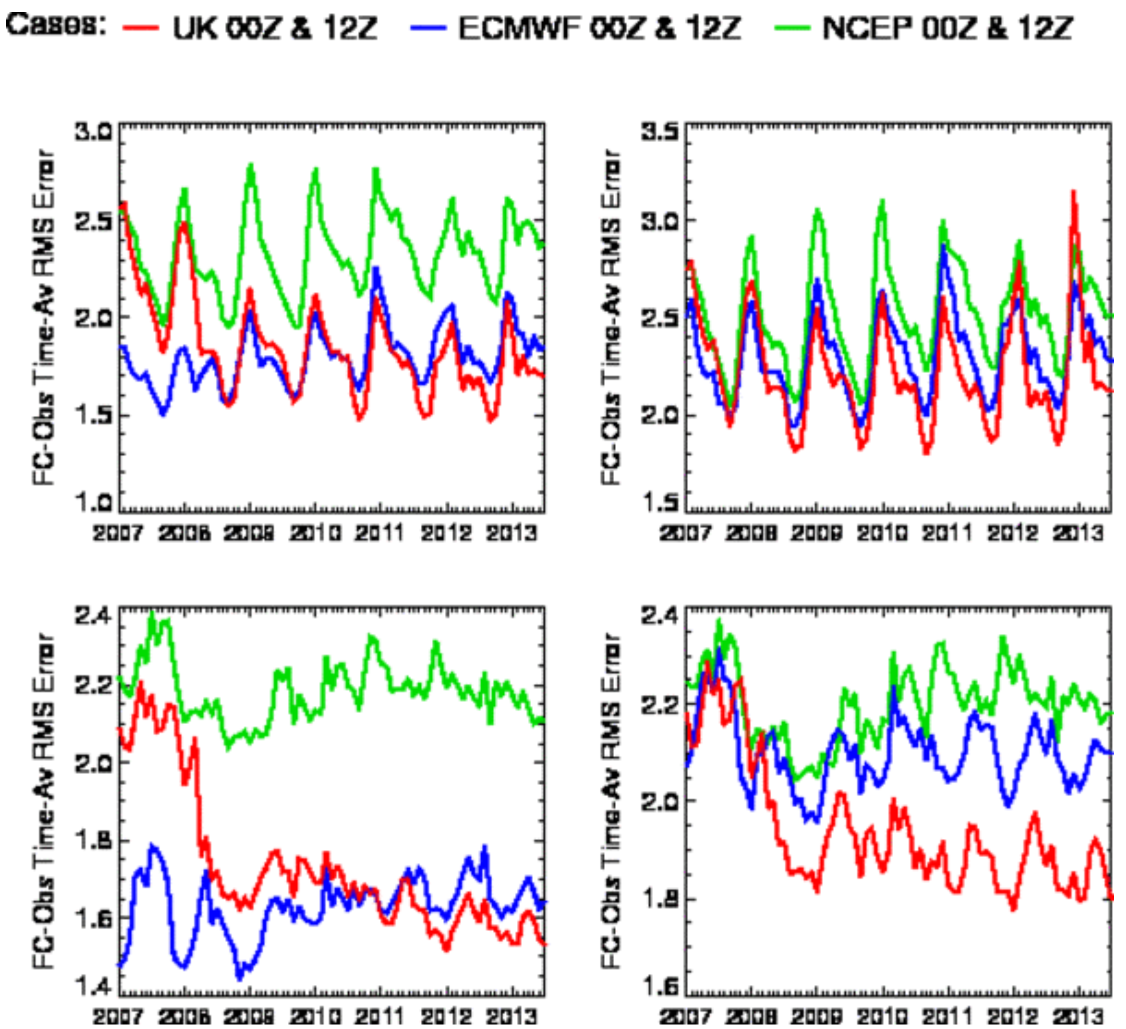
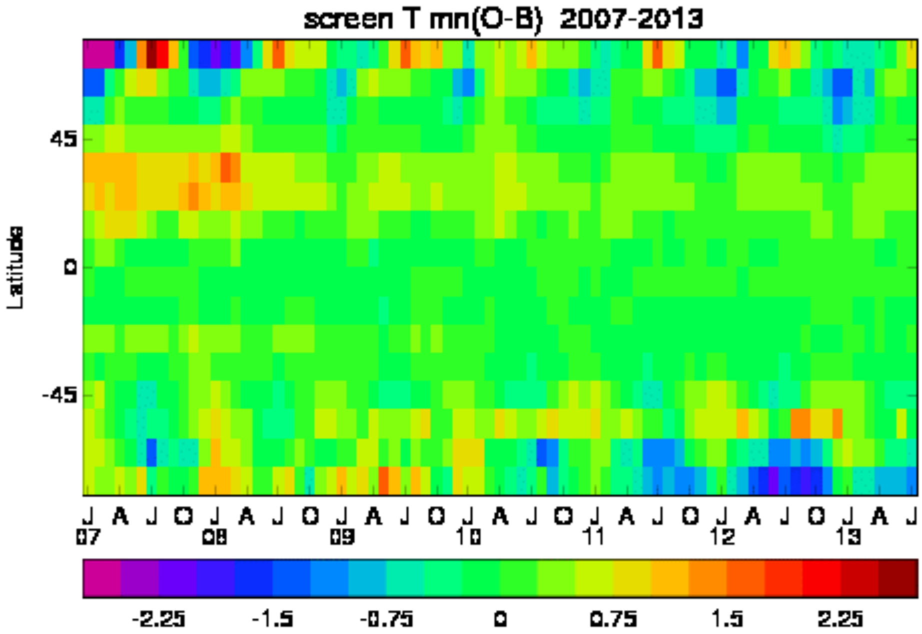


Figure 11. Verification of screen temperature, January 2007 – July 2013, Met Office in red. Top: 30-90°N, bottom 30°N-30°S; left T+0, right T+24 (courtesy of Rob Darvell). Verification is against Synop, Buoy and Ship observations but numerically Synops dominate.



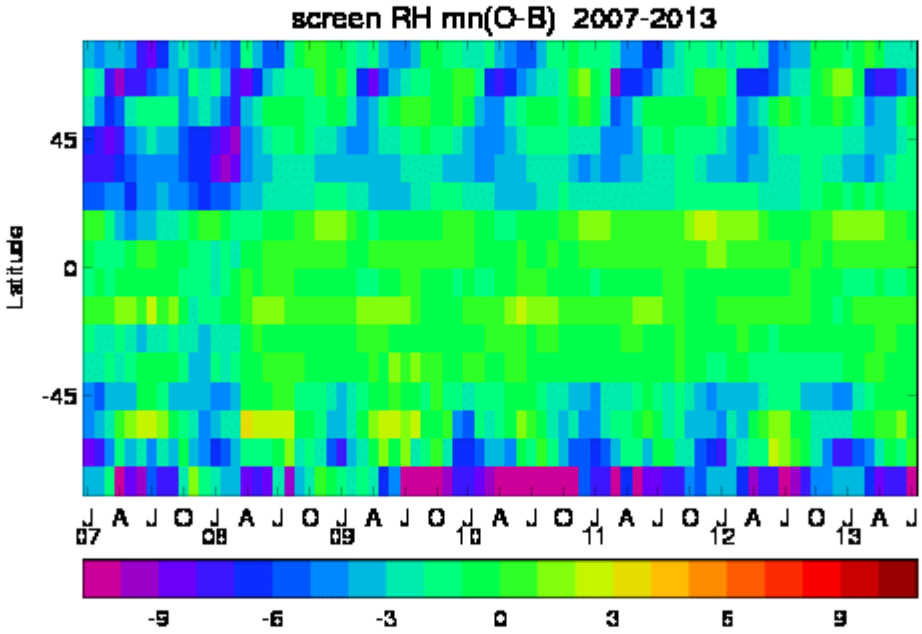


Figure 12. Temperature (°C) and RH (%) mean *o-b* differences for Synops, calculated monthly by 10° latitude bands (80°S-80°N).

Figure 12 shows T and RH *o-b* biases by latitude since 2007 – again the improvement in April 2008 can be seen. The model has a northward moving moist bias in northern hemisphere spring which seems to be connected to model snowmelt being too early by about three weeks. A change to a more sophisticated multi-layer snow model is planned. Seasonal high latitude temperature biases were slightly worse in 2012 than in previous years.

4.3. Forecast sensitivity to observations (FSO)

In recent years linearised forecast models and their adjoints have been used to estimate the impact of specific observation increments on forecasts. The results are specific to a particular forecast range, often 24 hours, and error norm. In the Canadian NWP system surface pressure data appears to be less important in 4D-Var than 3D-Var whereas surface temperature, humidity and wind data, whilst giving less impact than surface pressure, maintain similar values in 4D-Var (Rabier et al, 2007, figure 6). The reduction in sensitivity to surface pressure may be because 4D-Var can reconstruct some of the surface pressure field from satellite soundings.

Lorenc and Marriott (2013) produced FSO estimates for the Met Office system – their results used a moist norm at T+24 and a version of the NWP system just before Metar data were introduced (their figure 2 shows that the error of the linear forecast model and its adjoint is larger at low levels, this will affect the uncertainty of the estimates). Updated results for 20 January – 18 March 2012 are now available (courtesy of Richard Marriott), these suggest that Synop, Metar, buoy and ship contribute 8.4%, 3.6%, 4.3% and 0.6% respectively of the total impact (16.9% for surface data, compared to 10.0% for radiosondes and 9.2% for aircraft, most of the rest is from satellite data). Figure 13 shows a breakdown by surface observation type and variable. Surface pressure has the largest impact, with the impact from drifting buoys being particularly large compared to the number of reports - because many of them are in very data sparse areas such as the southern ocean. (The impact per Arctic drifting buoy temperature is even larger in this period.) Next in order of impact comes Synop and Metar temperature and then surface winds and RH over land. For a boreal summer period, 18 July - 22 August 2010 (Figure 14) land RH was much closer to land T in impact; it appears that humidity has a larger effect then because the moist norm uses specific humidity rather than RH and most of the stations are in the Northern hemisphere. Marine air temperature and RH have little impact: in most areas the air temperature is closely linked to the underlying sea surface temperature (SST) that is relatively slowly varying and also well observed by satellite as well as buoys and ships. FSO results for radiosondes by level (not shown) suggest that the impact per datum increases markedly going from 1000 hPa to 700 hPa – by about 2.5 times for T and RH, but by about 6 times for wind.

Maps of impact by variable (not shown) suggest that on balance T, RH and wind over Antarctica are slightly beneficial (this wasn't clear from earlier trials), but that impact from some stations/variables there is detrimental (this could be a sampling effect – discussed by Lorenc and Marriott – or it may be that tighter quality control would help). The impact maps suggested that there is a problem with the use of Metar winds (mainly from automated stations) over the Eastern USA – this drew attention to the wind biases seen in Figure 8. There also appear to be problems with the use of Synop winds in western Russia (possibly also a bias problem) and in areas near the Himalayas. In general the winds just poleward of 30° are beneficial – it is plausible that some of the tropical winds could be usefully assimilated.

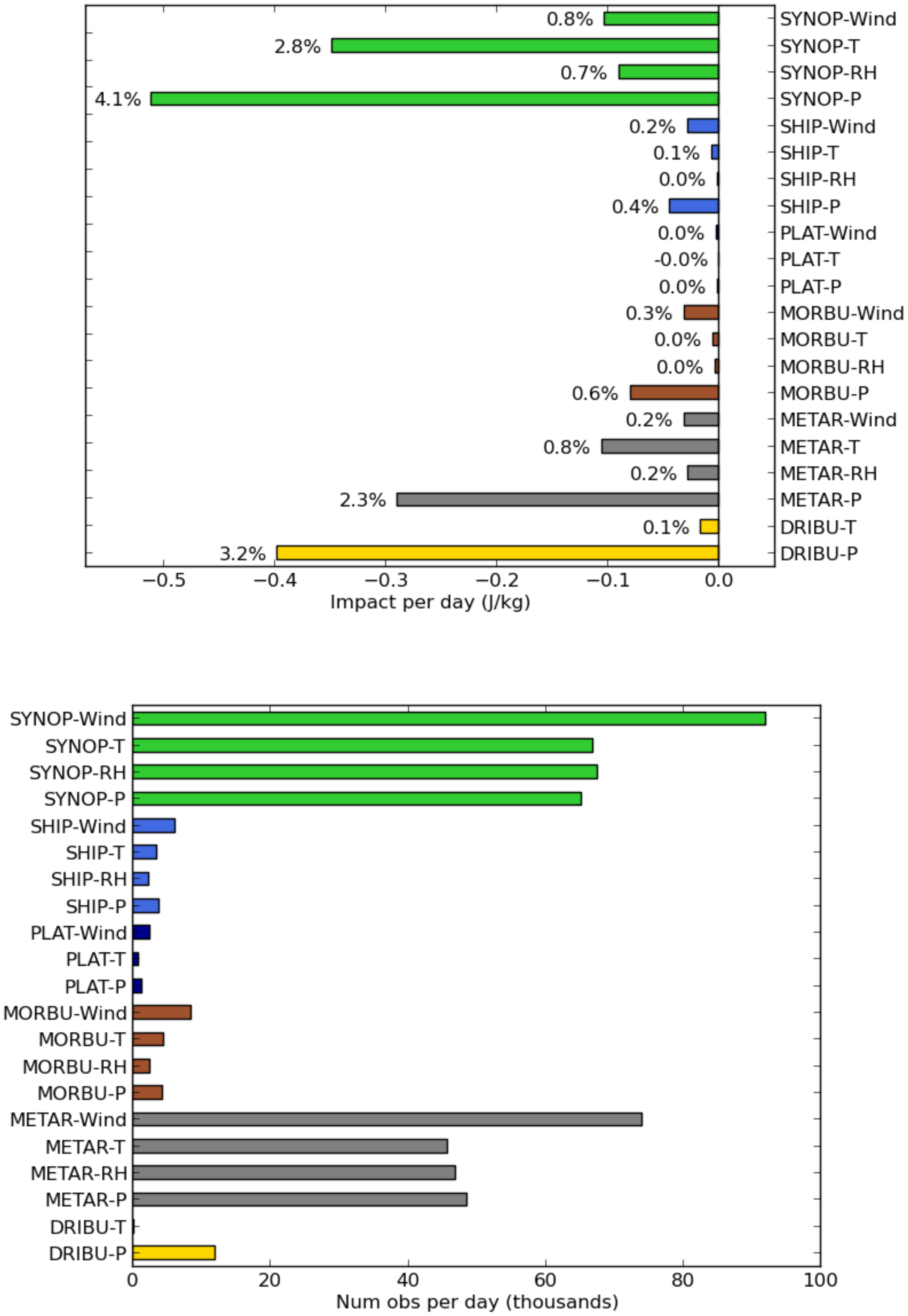


Figure 13. Adjoint based impact (negative values imply reduced error) of surface observations in our global forecast system and percentage of total impact (using a moist energy norm at T+24 for 30 January – 16 March 2013). Lower panel shows the number of observations assimilated per day (for wind the u and v components count as two

observations). Courtesy of Richard Marriott. (DRIBU – drifting buoy, MORBU – moored buoy, PLAT – platform/rig.)

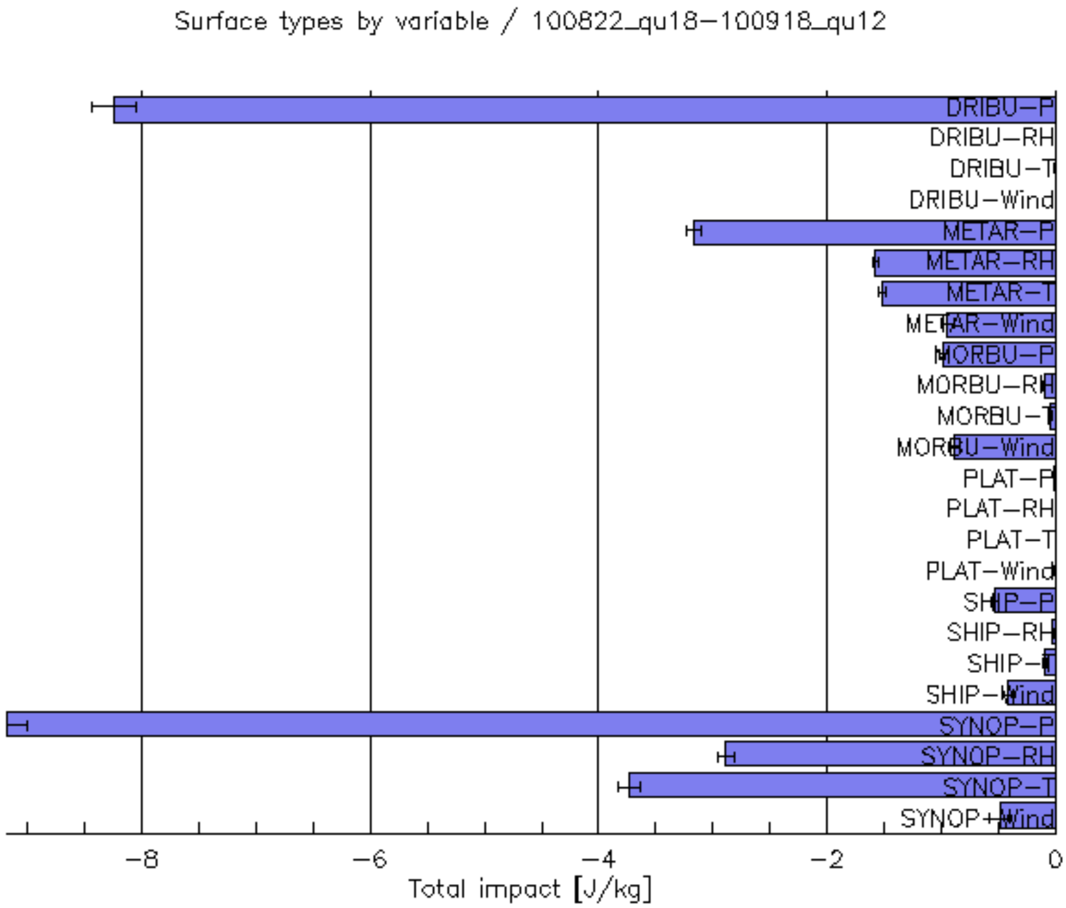


Figure 14. Adjoint based impact as in Figure 13 but for 22 August – 18 September 2010 and an earlier version of the NWP system.

5. Discussion and summary

5.1. Difficulties assimilating surface wind data

In a UK area forecasting system Macpherson et al (1996) found that "wind data at 10 m have very little impact beyond analysis time" and suggested that any projection onto the synoptic circulation above the boundary layer was minimal. Cansado and Navascues (2006) also found little impact from Synop wind data in a limited area model. Unlike the other variables considered wind background errors increase with height in the boundary layer (along with the wind speed which is approximately logarithmic in height at low levels). Thus analysis wind increments from a report at 10 m should perhaps increase above that height. However such extrapolation would magnify the impact of any observation errors and should be treated with caution. Ideally we would use wind reports from higher in the boundary layer: such as winds at 50 or 60 m from wind farms. The FSO results for radiosondes (last section) are consistent with wind usefulness increasing rapidly with height.

Apart from the vertical projection of wind increments there are issues of representativity (arguably worse for wind than for T and RH, especially near coasts) and wind biases.

The speed biases at night and in parts of the tropics need to be addressed before fully global assimilation of winds is possible. Very little assessment of surface winds appears to have been done on a global scale, although Vautard et al (2010) examined “stilling” in the Northern Hemisphere over recent decades. New et al (2002) appear to provide the only observation based climatology of wind speeds over land areas (Dunn et al, 2012, provide ungridded observations). To the best of the author’s knowledge none of the current generation of global reanalyses (including ERA-Interim) assimilate 10 m winds over land areas.

5.2. *Stable conditions*

Stable boundary layers pose particular problems for numerical modelling (Brown et al, 2008; ECMWF, 2012; Sandu et al, 2013), they are also challenging for data assimilation. Very stable conditions can occur where winds are weak and radiative heat loss is large (particularly on cloudless nights). There can be very large vertical gradients of temperature. This, and the possibility of errors in cloud cover, often mean particularly large *o-b* differences. Two questions arise: should we assimilate T, RH and wind at all under these circumstances, and if we do should we take account of the actual screen height (which can be between 1.25 and 2m)? Under very stable conditions there can be large differences between Tscreen and Tskin. Edwards et al, (2011) found that the MetUM underestimated these differences. Rejecting all nighttime data seems too drastic, but a rejection of particularly stable cases might be beneficial. Reen and Stauffer (2010) suggest that less vertical spreading of screen temperature increments would be appropriate under stable conditions. Hacker et al (2007) found that an Ensemble Kalman Filter could qualitatively produce this effect. In the UK area assimilation system Piccolo and Cullen (2011, 2012) have introduced an adaptive vertical grid which modifies the vertical correlations in stable conditions and improves fit to surface data.

5.3. *Verification issues*

The standard verification of surface fields is against Synops, ships and buoys (the Synops dominate numerically, Metars are not used). Stations more than 500 m above sea level are not used, this limit is applied to all variables although the motivation comes mainly from Pmsl. All pressure verification is performed on Pmsl, if Pstn was used (instead or in addition) then meaningful verification could be performed in mountainous areas (see section 2.2.1 and Ingleby, 1995). Blacklisted values and those that have failed quality control checks are not used for verification but we also have a “NoAssim” category which excludes values from assimilation but not from the verification. In this way tropical winds and winds from unresolved islands are included in the verification. The inclusion of island winds (despite their representativity issues) raises questions about the purpose of the verification – in principle one might use different exclusions for testing a revision to the current NWP system and for long-term monitoring of performance.

5.4. *Current/future work*

Comparison of surface *o-b* statistics between the Met Office and ECMWF (Ingleby et al, 2013b) indicates both similarities and differences between the errors of the two NWP systems. Comparison of data coverage revealed that some South African Synop reports weren't reaching the Met Office (they are now) and that some Metar reports weren't reaching ECMWF. Further checking for "missing" data is underway – new stations in data sparse regions would be especially welcome. WOW data (Section 2.1.4) will be monitored within the UK area forecasting system prior to possible assimilation. For both surface and radiosonde data the transition to binary (BUFR) data will require various processing changes and careful validation. Changes to the model dynamics and improved horizontal resolution (to ~17 km grid spacing) are being tested before implementation in early 2014 and give some improvements to surface variables and a reduction of surface wind speeds in some areas. The use of ensemble information in the data assimilation is likely to increase over the next few years.

For most stations the specified observation errors are larger than justified by the *o-b* statistics (tables 3-7). There was an attempt to reduce σ_o for pressure in late 2008 but this gave slightly worse forecast results – probably because the background errors were too large and the analysis is sensitive to the σ_o/σ_b ratio. The background errors have since been reduced so smaller σ_o should be tried again, but pressure σ_o should be a function of Δz (Figure 6) or perhaps pressure values with large $|\Delta z|$ should be rejected. As already discussed there are questions of whether to exclude nighttime winds in areas with speed biases, data in stable conditions or temperature in calm, sunny conditions when screens may overheat. Testing many small changes separately is rather expensive, testing them together usually means that their impacts cannot be separated. We cannot always assume that large rms *o-b* implies "bad observations" – the Arctic buoy temperatures have large *o-b* statistics (Ingleby, 2010) but are found to have a positive impact: they are valuable because they are in a very data sparse region.

5.5. Summary

Both in reality and in NWP models near-surface variables are strongly affected by soil, snow and surface properties, boundary layer structure and the diurnal cycle of radiative forcing modulated by cloud. Snow cover tends to increase temperature variability - so that temperature *o-b* statistics are largest at high latitudes. Operational surface measurements are not well documented in general – research measurements tend to be better documented but it often isn't clear how closely their performance compares with routine synoptic measurements. Satellite data are very important in NWP but in general cannot replace in situ measurements near the surface, especially over land (over the ocean satellite SST and scatterometer winds are important).

Assimilation of surface T/RH/wind data has a large impact on surface T and RH forecasts at short range and a small impact on Pmsl forecasts; there is a positive impact on surface wind analyses but little impact on the wind forecasts. Pressure assimilation impacts on Pmsl and upper air forecasts, but for surface temperature forecasts the temperature assimilation is much more important. The vertical profile of wind speed, and wind speed *o-b* biases especially at night, may be responsible for the lack of impact (or negative impact) of wind data. The poor agreement of forecast and reported winds in some regions is a particular concern. Precipitation in the forecast model is too high – hence forecast near-surface humidity is too high by a few percent. There is an additional peak in the humidity bias related to early snow melt in the northern extra-tropics. Despite these caveats the short range temperature and humidity forecasts generally verify very well.

In an NWP system observations are our link to reality, however a small proportion of observations have errors (of all conceivable kinds, including human typing errors). Using the data well needs significant attention to detail. This ranges from housekeeping issues (keeping position data up-to-date and complete), to scientific issues within the data assimilation system and interaction with model features. The quality of the observations is assessed before assimilating new types of data and ongoing monitoring is needed in order to reject any stations/variables with particular problems. Variables with marked biases from the model background need to be corrected or more often excluded (for example land wind speeds at model sea points compare badly with the background and are excluded). Height adjustment is very important for pressure and temperature and to a lesser extent for RH. Although the temperature lapse rate used works well on average it is less appropriate for stable conditions – stable conditions cause other problems both for modelling and data assimilation. At the Met Office the same computer code is used for processing observations for global and UK models which facilitates the propagation of improvements between the different systems.

Acknowledgements

The Met Office assimilation of surface data has benefitted over the years from work by Bruce Macpherson, Gareth Dow, Imtiaz Dharssi, Colin Parrett, Rick Rawlins, Adam Maycock and Mike Thurlow, data supply by Brian Barwell, Stan Kellett, Alice Ranford and Diana Rock and also from discussions with John Edwards and Adrian Lock in the boundary layer group and Mike Molyneux and David Moore of the observations section. Thanks also to Lars Isaksen for the chance to look at the ECMWF system.

6. References

Andersson E., E. Hólm, P. Bauer, A. Beljaars, G. A. Kelly, A. P., McNally, A. J. Simmons, J.-N. Thépaut, A. M. Tompkins, 2007: Analysis and forecast impact of the main humidity observing systems. Q. J. R. Meteorol. Soc. **133:** 1473-1485

Bell, S., Cornford, D. and Bastin, L., 2013: The state of automated amateur weather observations. Weather, **68:** 36-41. doi: 10.1002/wea.1980

Benjamin SG, Jamison BD, Moninger WR, Sahm SR, Schwartz BE, Schlatter TW, 2010: Relative Short-Range Forecast Impact from Aircraft, Profiler, Radiosonde, VAD, GPS-PW, METAR, and Mesonet Observations via the RUC Hourly Assimilation Cycle. Monthly Weather Review, **138:** 1319-1343 doi: http://dx.doi.org/10.1175/2009MWR3097.1

Brown, A., Beare, B., Edwards, J.M., Lock, A., Keogh S.J., Milton, S.F. and Walters, D.N., 2008: Upgrades to the boundary-layer scheme in the Met Office numerical weather prediction model. Boundary-Layer Meteorol, **128:** 117-132

Cansado, A. and B. Navascues, 2006: Usage of SYNOP 10 metres wind observations in HIRLAM. HIRLAM Newsletter # 50, July 2006 (via http://www.hirlam.org)

Cardinali, C., E. Andersson, P. Viterbo, J-N. Thépaut and Vasiljevic, D., 1994: Use of conventional surface observations in three-dimensional variational assimilation. ECMWF Technical Memorandum 207, Available from http://www.ecmwf.int/publications/library/do/references/list/14

Clark P.A., Harcourt, S.A., Macpherson, B., Mathison, C.T., Cusack, S., Naylor, M. 2008. Prediction of visibility and aerosol within the operational Met Office Unified Model. Part 1: model formulation and variational assimilation. Q. J. R. Meteorol. Soc., **134:** 1801-1816.

Clayton, A. M., Lorenc, A. C. and Barker, D. M., 2012: Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. Q. J. R. Meteorol. Soc.. doi: 10.1002/qj.2054

Compo GP, Whitaker JS, Sardeshmukh PD, Matsui N, Allan RJ, Yin X, Gleason Jr BE, Vose RS, Rutledge G, Bessemoulin P, Brönnimann S, Brunet M, Crouthamel RI, Grant AN, Groisman PY, Jones PD, Kruk MC, Kruger AC, Marshall GJ, Maugeri M, Mok HY, Nordli Ø, Ross TF, Trigo RM, Wang XL, Woodruff SD, Worley SJ. 2011. The Twentieth Century Reanalysis Project. Q. J. R. Meteorol. Soc. **137:** 1-28. DOI:10.1002/qj.776

Dharssi, I., P L Vidale, A Verhoef, B Macpherson, C Jones and M Best, 2008: New soil physical properties implemented in the Unified Model at PS18 Meteorology Research and Development Technical Report No. 528 Available from http://www.metoffice.gov.uk

Dharssi, I., Bovis, KJ., Macpherson, B. and C.P. Jones, 2011. Operational assimilation of ASCAT surface soil wetness. Hydrol. Earth System Sciences, **15:** 2729-2746. DOI: 10.5194/hess-15-2729-2011

Dow, G. and Macpherson B., 2013: Benefit of higher resolution analysis and observing systems in the UK Models. Forecasting Research Technical Report (draft)

Dumelow R and C Parrett, 2009: The impact of observation monitoring on Met Office NWP performance. Meteorology Research and Development Technical Report No. 533 Available from http://www.metoffice.gov.uk

Dunn, R. J. H., Willett, K. M., Thorne, P. W., Woolley, E. V., Durre, I., Dai, A., Parker, D. E., and Vose, R. S., 2012: HadISD: a quality-controlled global synoptic report database for selected variables at long-term stations from 1973-2011, Clim. Past, **8:** 1649-1679, doi:10.5194/cp-8-1649-2012

ECMWF, 2012: Proceedings of ECMWF Workshop on Diurnal cycles and the stable boundary layer, 7-10 November 2011. Available from http://www.ecmwf.int/publications/library/do/references/list/201111

Edwards J. M. 2009: Radiative Processes in the Stable Boundary Layer: Part II. The Development of the Nocturnal Boundary Layer. Boundary-Layer Meteorol. **131:** 127-146 doi 10.1007/s10546-009-9363-9

Edwards JM, 2012: Radiation and turbulence in the stable boundary layer, Proceedings of ECMWF Workshop on Diurnal cycles and the stable boundary layer, 7-10 November 2011. http://www.ecmwf.int/publications/library/do/references/list/201111

Edwards JM, McGregor JR, Bush MR, Bornemann FJ. 2011. Assessment of numerical weather forecasts against observations from Cardington: seasonal diurnal cycles of screen-level and surface temperatures and surface fluxes. Q. J. R. Meteorol. Soc. **137:** 656–672. DOI:10.1002/qj.742

Flowerdew, J. and N. E. Bowler, 2013: On-line calibration of the vertical distribution of ensemble spread. Q. J. R. Meteorol. Soc. doi: 10.1002/qj.2072

Gandin, L. S., L. L. Morone, W. G. Collins, 1993: Two Years of Operational Comprehensive Hydrostatic Quality Control at the National Meteorological Center Weather and Forecasting **8:** 57-72

Genthon, C., D. Six, V. Favier, M. Lazzara, and L. Keller, 2011: Atmospheric temperature measurement biases on the Antarctic Plateau. J. Atmos. Oceanic Technol., **28:** 1598-1605. doi: http://dx.doi.org/10.1175/JTECH-D-11-00095.1

Hacker JP, Anderson JL, Pagowski M, 2007: Improved vertical covariance estimates for ensemble-filter assimilation of near-surface observations. Mon. Wea. Rev. **135:** 1021-1036

Healy, S. B. 2013: Surface pressure information retrieved from GPS radio occultation measurements. Q. J. R. Meteorol. Soc. doi: 10.1002/qj.2090

Horlacher V., S Osborne, JD Price, 2012: Comparison of two closely located meteorological measurement sites and consequences for their areal representivity Boundary-Layer Meteorol., **142:** 469-493 doi: 10.1007/s10546-011-9684-3

Howard, T. and Clark, P., 2003: Improvement to the Nimrod Wind Nowcasting Scheme over High Ground. Forecasting Research Technical Report 406. Available from http://www.metoffice.gov.uk

Howard, T. and Clark, P., 2007: Correction and downscaling of NWP wind speed forecasts. Meteorological Applications, **14:** 105-116, doi: 10.1002/met.12

Hubbard, K. G., X. Lin, C. B. Baker and B. Sun, 2004: Air temperature comparison between the MMTS and the USCRN temperature systems. Journal of Atmospheric and Oceanic Technology **21:** 1590-1597

Hobby, M., M. Gascoyne, J. Marsham, M. Bart, C. Allen, S. Engelstaedter, D. Fadel, A. Gandega, R. Lane, J. McQuaid, B. Ouchene, A. Ouladichir, D. Parker, P. Rosenberg, M. Ferroudj, A. Saci, F. Seddik, M. Todd, D. Walker, and R. Washington, 2012: The Fennec Automatic Weather Station (AWS) Network: monitoring the Saharan Climate System. J. Atmos. Oceanic Technol. **30:** 709-724 doi:10.1175/JTECH-D-12-00037.1

ICAO, 2006: Manual on Automatic Meteorological Observing Systems at Aerodromes, first Ed, ICAO Doc 9837 AN/454

Ingleby, N. B., 1995: Assimilation of station level pressure and errors in station height. Weather and Forecasting, **10:** 172-182

Ingleby, N. B., 2001: The Statistical Structure of Forecast Errors and its Representation in The Met. Office Global 3-D Variational Data Assimilation Scheme. Q. J. R. Meteorol. Soc. **127:** 209-231

Ingleby, B. 2010: Factors affecting ship and buoy data quality: a data assimilation perspective. Journal of Atmospheric and Oceanic Technology, **27:** 1476-1489. doi: 10.1175/2010JTECHA1421.1

Ingleby, N. B., A. C. Lorenc, K. Ngan, F. Rawlins and D. R. Jackson, 2012: Improved variational analyses using a nonlinear humidity control variable. Q. J. R. Meteorol. Soc. doi:10.1002/qj.2073.

Ingleby, B., D. Moore, C. Sloan, R. Dunn, 2013a: Evolution and accuracy of surface humidity reports. J. Atmos. Oceanic Technol., accepted

Ingleby, B., L. Isaksen and M. Dahoui, 2013b: Comparison of short range Met Office and ECMWF forecasts with surface observations. 4th WGNE workshop on systematic errors in weather and climate models, available from http://www.metoffice.gov.uk/conference/wgne2013

Janjic, T., SE Cohn, 2006: Treatment of observation error due to unresolved scales in atmospheric data assimilation. Mon. Wea. Rev. **134:** 2900-2915 doi: 10.1175/MWR3229.1

Kållberg, P., 1998: Aspects of the re-analysed climate, ERA-15 Project Report Series No. 2. http://www.ecmwf.int/publications/library/do/references/list/191

Kristensen, L., 1998: Cup anemometer behavior in turbulent environments. J. Atmos. Oceanic Technol., **15:** 5-17.

Kumamoto, M., M. Otsuka, T. Sakai, T. Hamagami, H. Kawamura, T. Aoshima, and F. Fujibe, 2013: Field experiment on the effects of a nearby asphalt road on temperature measurement. SOLA, **9:** 56-59, doi:10.2151/sola.2013-013.

Lapworth A, 2003: Factors determining the decrease in surface wind speed following the evening transition. Q. J. R. Meteorol. Soc. **129:** 1945-1968 doi 10.1256/qj.02.163

Laroche, S. and R. Sarrazin. 2010. Impact study with observations assimilated over North America and the North Pacific Ocean on the MSC global forecast system. Part I: contribution of radiosonde, aircraft and satellite data. ATMOSPHERE-OCEAN, **48:** 10–25.

Lazzara, M. A., G. A. Weidner, L. M. Keller, J. E. Thom, J. J. Cassano, 2012: Antarctic Automatic Weather Station Program: 30 Years of Polar Observation. Bull. Amer. Meteor. Soc., **93:** 1519-1537. doi: http://dx.doi.org/10.1175/BAMS-D-11-00015.1

Li, Z.-L., Tang, B.-H., Wu, H., Ren, H., Yan, G., Wan, Z., Trigo, I. F., and Sobrino, J. A.2013: Satellite-derived land surface temperature: Current status and perspectives, Remote Sens. Environ., **131:** 14-37, doi:10.1016/j.rse.2012.12.008

Lin X, Hubbard KG, Baker CB, 2005: Surface air temperature records biased by snow-covered surface. International Journal of Climatology, **25:** 1223-1236

Lin, X., and K. G. Hubbard, 2008: What are daily maximum and minimum temperatures in observed climatology? International Journal of Climatology **28:** 283-294

Lock, A. P., A. R. Brown, M. R. Bush, G. M. Martin, R. N. B. Smith, 2000: A New Boundary Layer Mixing Scheme. Part I: Scheme Description and Single-Column Model Tests. Mon. Wea. Rev., **128:** 3187–3199.

Lock, A., J. Edwards, 2012: The Parametrization of Boundary Layer Processes. Unified Model Documentation Paper 24. Available from Met Office.

Lorenc, A. C., S. P. Ballard, R. S. Bell, N. B. Ingleby, P. L. F. Andrews, D. M. Barker, J. R. Bray, A. M. Clayton, T. Dalby, D. Li, T. J. Payne and F. W. Saunders, 2000: The Met. Office global three-dimensional variational data assimilation scheme, Q. J. R. Meteorol. Soc. **126:** 2991-3012

Lorenc, A.C. and Hammon, O., 1988: Objective quality control of observations using Bayesian methods. Theory, and a practical implementation. Q. J. R. Meteorol. Soc., **114:** 515-543

Lorenc, A. C. and Marriott, R. T., 2013: Forecast sensitivity to observations in the Met Office Global numerical weather prediction system. Q.J.R. Meteorol. Soc.. doi: 10.1002/qj.2122

Macpherson B., Wright B.W., Hand W.H. and Maycock A.J. 1996: The impact of MOPS moisture data in the UK Meteorological Office data assimilation scheme. Mon. Weath. Rev. **124:** 1746-1766.

Mahmood R, Foster SA, Logan D, 2006: The GeoProfile metadata, exposure of instruments, and measurement bias in climatic record revisited, I J Climatol, **26:** 1091-1124

Mahoney, W. P., J. M. O'Sullivan, 2013: Realizing the Potential of Vehicle-Based Observations. Bull. Amer. Meteor. Soc., **94:** 1007-1018.
doi: http://dx.doi.org/10.1175/BAMS-D-12-00044.1

Makkonen, L., P. Lehtonen, L. Helle, 2001: Anemometry in Icing Conditions. J. Atmos. Oceanic Technol., **18:** 1457–1469.

McCutchan M. H. 1983: Comparing temperature and humidity on a mountain slope and in the free air nearby. Monthly Weather Review **111:** 836-845

McCutchan M. H. and D. G. Fox, 1986: Effect of Elevation and Aspect on Wind, Temperature and Humidity. Journal of Applied Meteorology **25:** 1996-2013

Milton, S.F., and P. Earnshaw, 2007: Evaluation of surface water and energy cycles in the Met Office global NWP model using CEOP data. J. Meteorol. Soc. Japan, **85A:** 43-72

Nakamura, R. and L.Mahrt, 2005: Air temperature measurement errors in naturally ventilated radiation shields. Journal of Atmospheric and Oceanic Technology, **22:** 1046-1058

National Academy of Sciences, 2009: Observing Weather and Climate from the Ground Up: A Nationwide Network of Networks, Available from http://dels.nas.edu/Report/Observing-Weather-Climate-from/12540

New M, D Lister, M Hulme, I Makin - Climate Research, 2002: A high-resolution data set of surface climate over global land areas. Climate Research, 2002. **21:** 1-25

Painter, H. E., 1977. An analysis of the differences between dry-bulb temperatures obtained from an aspirated psychrometer and those from a naturally ventilated large thermometer screen at Kew Observatory. Meteorological Office unpublished report (National Meteorological Library, Exeter).

Pepin N.C. and D. J. Seidel, 2005: A global comparison of surface and free-air temperatures at high elevations. Journal of Geophysical Research-Atmospheres **110** (D3), D03104

Piccolo, C. and Cullen, M. 2011: Adaptive mesh method in the Met Office variational data assimilation system. Q.J.R. Meteorol. Soc., **137:** 631-640. doi: 10.1002/qj.80

Piccolo, C. and Cullen, M. 2012: A new implementation of the adaptive mesh transform in the Met Office 3D-Var System. Q.J.R. Meteorol. Soc., **138:** 1560-1570. doi: 10.1002/qj.1880

Pu, Z., Zhang, H., Anderson, J. 2013: Ensemble Kalman filter assimilation of near-surface observations over complex terrain: comparison with 3DVAR for short-range forecasts. Tellus A, **65:**

Qin, Z.-K., X. Zou, G. Li, and X.-L. Ma 2010: Quality control of surface station temperature data with non-Gaussian observation-minus-background distributions, J. Geophys. Res., **115:** D16312, doi:10.1029/2009JD013695.

Rabier F., P. Gauthier, C. Cardinali, R. Langland, M. Tsyrulnikov, A. Lorenc, R. Gelaro, P. Steinle, K. Koizumi 2007: An update on THORPEX-related research in Data Assimilation and Observing Strategies. Nonlin. Processes Geophys., **15:** 81-94.

Rawlins, F., S. P. Ballard, K. J. Bovis, A. M. Clayton, Dingmin Li, G. W. Inverarity, A. C. Lorenc, T. J. Payne, 2007: The Met Office Global 4D-Var Scheme. Q. J. R. Meteorol. Soc., **133:** 347-362

Reen B.P., and D.R. Stauffer, 2010: Data Assimilation Strategies in the Planetary Boundary Layer. Boundary-Layer Meteorol, **137:** 237-269

Rolland, C. 2003: Spatial and seasonal variations of air temperature lapse rates in Alpine regions, J. Climate, **16:** 1032-1046

Sandu, I., A. Beljaars, P. Bechtold, T. Mauritsen, and G. Balsamo, 2013: Why is it so difficult to represent stably stratified conditions in numerical weather prediction (NWP) models?, J. Adv. Model. Earth Syst., **5:** doi:10.1002/jame.20013.

Sheridan P., S. Smith, A. Brown and S. Vosper, 2010: A simple height-based correction for temperature downscaling in complex terrain. Meteorological Applications. **17:** 329-339, doi: 10.1002/met.177

Simmons, A. J., K. M. Willett, P. D. Jones, P. W. Thorne, and D. P. Dee, 2010: Low-frequency variations in surface atmospheric humidity, temperature, and precipitation: Inferences from reanalyses and monthly gridded observational data sets, J. Geophys. Res., **115:** D01110, doi:10.1029/2009JD012442

Sloan, C. and M. R. Clark, 2012: A comparison of three Met Office wind observing systems. Atmosph. Sci. Lett., **13:** 283-288. doi: 10.1002/asl.396

Tennant, W. and S Beare, 2013: New schemes to perturb sea-surface temperature and soil moisture content in MOGREPS. Q.J.R. Meteorol. Soc. doi: 10.1002/qj.2202

Turton, J., C. Pethica, 2010: Assessment of a New Anemometry System for the Met Office's Moored Buoy Network. J. Atmos. Oceanic Technol., **27:** 2031-2038. doi: 10.1175/2010JTECHA1475.1

Tyndall, D.P., J.D. Horel and M.S.F.V. de Pondeca, 2010: Sensitivity of Surface Air Temperature Analyses to Background and Observation Errors. Weather and Forecasting **25:** 852-865

Vasiljevic, D. 2006: Handling biases in surface pressure (Ps) observations in data assimilation. ECMWF/EUMETSAT NWP-SAF Workshop on Bias estimation in data assimilation, 8-11 November 2005. Available from http://www.ecmwf.int/publications/library/do/references/list/22022006

Vautard R, J Cattiaux, P Yiou, J-N Thépaut, P Ciais, 2010: Northern Hemisphere atmospheric stilling partly attributed to an increase in surface roughness. Nature Geoscience, **3:** 756-761, doi:10.1038/ngeo979

Waller J. A., S. L. Dance, A. S. Lawless, N. K. Nichols, J. R. Eyre, 2013: Representativity error for temperature and humidity using the Met Office high resolution model. Q. J. R. Meteorol. Soc. Accepted

Wood,N. and Mason,P.J., 1993: The pressure force induced by neutral turbulent flow over hills. Q. J. R. Meterol. Soc., **119,** 1233-1297.

WMO, 2010: *Guide to meteorological instruments and methods of observation*, Geneva, World Meteorological Organisation (2008 edition, updated in 2010) Available at http://www.wmo.int/pages/prog/www/IMOP/CIMO-Guide.html

WMO, 2011: *WMO Publication No. 306 - Manual on codes. Volume I.1 and Volume I.2* Available at http://www.wmo.int/pages/prog/www/WMOCodes.html

Appendix 1 Pressure processing

Metar processing

Metar reporting is regulated by the International Civil Aviation Organisation (ICAO). At airports the measured pressure is used to calculate QFE and QNH:
QFE is the pressure reduced to official airfield altitude (a small adjustment).
QNH (calculated from QFE) is the pressure reduced to mean sea level, using the ICAO standard profile of the atmosphere.
Officially (ICAO, 2006) QNH should be rounded down (for reasons of aircraft safety) to the nearest whole hPa for the Metar report. Some countries, including the USA and Canada, report values in hundredths of an inch of mercury (~ 0.33864 hPa). Such reports make up about 40% of total Metar reports. For values reported in whole hPa we add 0.5 hPa to avoid a systematic bias, the other values are used unadjusted.

Pressures rounded to whole hPa have their observation error variance increased to take account of this. If the error is ε+r where r is sampled from a uniform distribution [-b, b] independent of ε then $\left\langle (\varepsilon + r)^2 \right\rangle = \left\langle \varepsilon^2 \right\rangle + \frac{1}{2b}\int_{-b}^{b} r^2 dr = \sigma_o^2 + \frac{b^2}{3}$. $b^2/3 = 0.08333$ for b=0.5 hPa. If σ_o is 1.0 (0.5) then the modified value is 1.041 (0.577). The same error is used for the different pressure quantities although P_* errors will be smaller than Pmsl errors for high level stations.

There are various equations in use worldwide for calculating QNH from QFE, but they should all give very similar results to that of ICAO (2006):

$$H = 44330.77 - 11880.32 QFE^{0.190263} \tag{1.1}$$

$$QNH = 1013.25\left\{1 - 0.0065(H - HA)/288.15\right\}^{5.25588} \tag{1.2}$$

H is the equivalent altitude in the ICAO standard atmosphere (in m). In our processing these equations are inverted:

$$H = \frac{288.15\left\{1 - (QNH/1013.25)^{1/5.25588}\right\}}{0.0065} + HA \tag{1.3}$$

$$QFE = \left\{\frac{44330.77 - H}{11880.32}\right\}^{1\ 0.190263} \tag{1.4}$$

Up to this point all pressures are in hPa. We set $p_{stn} = 100 QFE$ in Pa and $z_{stn} = HA$.

Conversion of station level pressure to model surface height

The determination of background pressure at the observation location involves both horizontal interpolation and vertical adjustment, in regions of steep orography care is needed with the interaction between them. The basic method is as described by Ingleby (1995) but the details below are due to Berney (1999, pers. comm.). Assuming a constant lapse rate L=-dT/dz (taken as 0.0065 K m^{-1}) we integrate the hydrostatic equation, dp/dz = -pg/RT, between p_0 and p_1 giving

$$p_1/p_o = (T_1/T_0)^{g/RL} \tag{1.5}$$

R is the gas constant per mole and *g* the acceleration due to gravity. Rather than use the model temperature near the surface (which is subject to diurnal/local variations) we start from a model virtual temperature T_{2000} about 2000 m above the model surface (at pressure p_{2000}) and then derive T_0 at p_0=p_* using $T_o = T_{2000}\left(p_p / p_{2000}\right)^{RL/g}$. Using $T_1 = T_0 + L(z_o - z_1)$ equation (1.5) can be rearranged to give

$$A - z_1 - Bp_1^{RL/g} = 0$$
$$where \quad A = z_0 + T_o \; L \quad and \quad B = \frac{T_o}{Lp_0^{RL/g}} \qquad (1.6)$$

A and B are calculated at each model grid point, and bilinearly interpolated in the horizontal to the observation position. The result is equivalent to the linear interpolation of the geopotential height of constant pressure surfaces, and preserves the assumed lapse rate at all points. This gives p_1 - background pressure at the observation height (z_{stn}, or 0 if Pmsl is used) which is compared to p_{stn}. In practice we then adjust p_{stn} to the model height using $p_{*ob} = p_{stn}\left(p_0 / p_1\right)$.

The default is to use Pstn but where the monitoring shows that the Pmsl quality is clearly better Pmsl is used. Biases of over 1.5 hPa in magnitude are corrected using monthly *o-b* statistics[3]. Because surface pressure is a reference for the NWP system we do not want to correct all the surface pressures. Using March 2013 data 825 Synop stations have a pressure or height correction applied (height corrections are applied to the few cases where the uncorrected bias is more than 15 hPa). For Metar, Ship/Buoy and Mobile Synop stations the numbers are 123, 351 and 74 respectively. The proportion for Metars is quite low - probably because the station pressure is a vital aid for landing aircraft safely. Some of the largest pressure corrections are over Antartica, and the monthly monitoring, including these corrections, has most impact in the Southern Hemisphere (Dumelow and Parrett, 2009).

In November 2009 there were changes to the processing of Synop pressure data to a) use more Pmsl values where Pmsl statistics are clearly better than Pstn statistics (1240 stations as at March 2013) and b) to bias correct Pmsl and Pstn separately (rather than to correct P_* after it has been calculated). a) gave a clear improvement, b) did not - although it will allow corrected Pmsl to be used in verification against observations. P_* corrections are to some extent dependent on the model orography and hence the model horizontal resolution - the Pmsl and Pstn corrections should be more independent of the model.

[3] ECMWF have an automated pressure bias correction algorithm which uses a running archive of about seven day's data (Vasiljevic, 2006).