University of Reading

Comparing diagnosed observation uncertainties with independent estimates: a case study using aircraft-based observations and a convection-permitting data assimilation system

Article

Published Version

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1002/asl.1029

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

RESEARCH ARTICLE

Comparing diagnosed observation uncertainties with independent estimates: A case study using aircraft-based observations and a convection-permitting data assimilation system

Andrew K. Mirza[1] | Sarah L. Dance[1] | Gabriel G. Rooney[3] |
David Simonin[2] | Edmund K. Stone[3] | Joanne A. Waller[1,2]

[1]School of Mathematical, Physical and Computational Sciences, University of Reading, Reading, United Kingdom

[2]MetOffice@Reading, University of Reading, Reading, United Kingdom

[3]Met Office, Exeter, United Kingdom

Correspondence
Andrew K. Mirza, School of Mathematical, Physical and Computational Sciences, University of Reading, Reading, RG6 6BB, United Kingdom.
Email: akmirza@mail.com

Funding information
Engineering and Physical Sciences Research Council, DARE project, Grant/Award Number: EP/P002331/1; Natural Environment Research Council, Flooding from Intense Rainfall programme, Grant/Award Number: NE/K008900/1

Abstract

Aircraft can report in situ observations of the ambient temperature by using aircraft meteorological data relay (AMDAR) or these can be derived using mode-select enhanced tracking data (Mode-S EHS). These observations may be assimilated into numerical weather prediction models to improve the initial conditions for forecasts. The assimilation process weights the observation according to the expected uncertainty in its measurement and representation. The goal of this paper is to compare observation uncertainties diagnosed from data assimilation statistics with independent estimates. To quantify these independent estimates, we use metrological comparisons, made with in-situ research-grade instruments, as well as previous studies using collocation methods between aircraft (mostly AMDAR reports) and other observing systems such as radiosondes. In this study we diagnose a new estimate of the vertical structure of the uncertainty variances using observation-minus-background and observation-minus-analysis statistics from a Met Office limited area three-dimensional variational data assimilation system (3 km horizontal grid-length, 3-hourly cycle). This approach for uncertainty estimation is simple to compute but has several limitations. Nevertheless, the resulting diagnosed variances have a vertical structure that is like that provided by the independent estimates of uncertainty. This provides confidence in the uncertainty estimation method, and in the diagnosed uncertainty estimates themselves. In the future our methodology, along with other results, could provide ways to estimate the uncertainty for the assimilation of aircraft-based temperature observations.

KEYWORDS

aircraft-based observations, AMDAR, data assimilation, estimation of observation uncertainty, meteorological instruments, Mode-S, observational data analysis

1 | INTRODUCTION

In data assimilation, observations and prior model forecasts are combined, taking account of their relative uncertainties, to provide initial conditions for numerical weather prediction (NWP). Improved specification of observation uncertainties has been shown to improve analysis accuracy and forecast skill (e.g., Stewart *et al.*, 2008, 2013; Weston *et al.*, 2014; Bormann *et al.*, 2016; Campbell *et al.*, 2017; Simonin *et al.*, 2019). However, observation uncertainty is difficult to measure and can only be estimated by gathering appropriate statistics. A simple approach for estimating the observation error covariance, known as the Desroziers *et al.* (2005) diagnostic, has become popular due to its ease of use. This method uses samples of observation-model departures routinely output from the data assimilation system and has been applied to a number of observation types, such as Doppler radial winds (Waller *et al.*, 2016a, 2019), satellite radiances, (e.g., Stewart *et al.*, 2014; Waller *et al.*, 2016b) smoothed aircraft-derived observations (Lange and Janjic, 2016), and surface observations (Tavolato and Isaksen, 2015). It is known that this diagnostic approach relies on unrealistic assumptions, and provides only an approximation to the observation error covariance (Desroziers *et al.*, 2005; Ménard, 2016; Waller *et al.*, 2016; Waller *et al.*, 2017; Bathmann, 2018). However, there have only been limited studies comparing the uncertainty estimates computed using the Desroziers *et al.* (2005) diagnostic with independent estimates of the observation error statistics (e.g., contributions from measurement error, quality control processing, observation error and error due to unresolved scales Chun *et al.*, 2015).

Numerous studies have shown that aircraft-based observations are a valuable source of information for NWP (e.g., Cardinali *et al.*, 2003; Lorenc and Marriott, 2014; Lange and Janjic, 2016; Petersen, 2016; James and Benjamin, 2017; Ingleby *et al.*, 2019; Mirza *et al.*, 2019). Aircraft temperature reports are the processed reports from the aircraft's sensor and avionics. Temperature reports may also be derived from processed Mode Select (Mode-S) reports obtained from suitably configured secondary surveillance radar. Processed reports are mostly obtained from aircraft that participate in the Aircraft Meteorological Data Relay (AMDAR) program (Stickland and Grooters, 2005; WMO, 2017), or could be obtained from Mode-S Meteorological Routine Air Reports (MRAR) (Strajnar, 2012). Derived temperature reports are obtained from Mode-S enhanced surveillance (EHS) reports of the aircraft's true airspeed and Mach number (de Haan, 2011; Stone and Kitchen, 2015; Mirza *et al.*, 2016). Mode-S EHS reports are much more frequent than either AMDAR or Mode-S MRAR (de Haan, 2011; Strajnar, 2012; Stone and Pearce, 2016; Mirza, 2017). However, the derived temperature reports suffer from large uncertainty, especially at low altitudes, due to the low precision in the reported Mach number. This source of uncertainty has been studied using a metrological approach by Mirza *et al.* (2016). De Haan (2011) and Mirza (2017) have shown that the uncertainty can be reduced by application of smoothing filters. These smoothed data are used in NWP at several centres. In the literature Mode-S EHS is used synonymously for unsmoothed and smoothed derived temperatures (de Haan, 2011; de Haan and Stoffelen, 2012; Lange and Janjic, 2016). However, in this paper, we aim to diagnose the error for the *unsmoothed derived temperature*, hereafter called the Mach temperature, T_{MACH} (Mirza *et al.*, 2016, 2019). We choose to focus on the analysis of T_{MACH} as uncertainties in aircraft winds, reported by AMDAR and derived from Mode-S EHS, have already been subject to more extensive studies, summarized by Mirza *et al.* (2016, section 2).

In this paper, we use the Desroziers *et al.* (2005) diagnostic to compute new estimates of observation error variances for AMDAR temperature and T_{MACH} reports. Unlike Lange and Janjic (2016), we compare the diagnosed estimates with two types of independent estimates of observation error variances computed by (a) collocation with other observations and (b) metrological analysis. In Section 2, we define our terms and outline the methods used to obtain temperature reports from aircraft; compare empirical and assumed estimates of observation uncertainties; and we describe briefly the Desroziers *et al.* (2005) diagnostic which is widely used to estimate observation error covariance matrices from observation-model departures. In Section 3, we describe the experimental methods using the Met Office limited area NWP model for the United Kingdom (UKV) to diagnose the observation error variances. Section 4 shows the newly diagnosed uncertainties in comparison with empirical estimates. We summarize our findings in Section 5 and note that these results provide new confidence in variances diagnosed using the Desroziers *et al.* (2005) technique.

2 | ESTIMATING OBSERVATION UNCERTAINTY OF TEMPERATURE REPORTS FROM AIRCRAFT

Numerous studies have computed observation uncertainty estimates for aircraft-based observations (see review by Mirza *et al.*, 2016); in this section we outline these methods and their limitations for estimating observation uncertainty.

To distinguish between various methods used to estimate observation variances we will refer to assumed, empirical and diagnosed estimates; these are defined

below in Sections 2.1, 2.2 and 2.3, respectively. The main sources of error for these methods may be due to one or more of the following: instrumental error, effects due to atmospheric boundary layer turbulence; for aircraft at low altitudes, the combined effects of low airspeed, airframe configuration and aircraft manoeuvres (Drüe et al., 2008); in the case of Mode-S EHS, the precision of its reports (de Haan, 2011; Mirza et al., 2016); and errors arising from the NWP system, for example, pre-processing or quality-control, the observation operator, and unresolved scales and processes (Janjic et al., 2018).

2.1 | Assumed estimates

Assumed estimates are an idealized observation uncertainty used in the NWP data assimilation system. They are based on the further analysis of error estimates, using linear or polynomial regression methods, and/or use expert judgement to produce a smoothed estimate. These assumed estimates may also be artificially inflated or deflated to take account of other sources of error in the observation and/or NWP system which may be difficult or too costly to quantify precisely but are assumed to exist, for example, the measurement of small scale variability by the observing system which is not resolvable within in the NWP system. Ideally, these assumed estimates should be reviewed periodically to take account of developments in the observation and NWP systems, although this may not always occur in practice. Figure 1 shows the vertical profiles of assumed estimates of observation uncertainty standard deviation for AMDAR temperature reports as used for a number of NWP data assimilation systems: Consortium for Small-Scale Modeling Kilometre-scale Ensemble Data Assimilation (COSMO-KENDA) (Schraff et al., 2016, table 1, p. 1457); Met Office limited-area, high-resolution, convection-permitting NWP
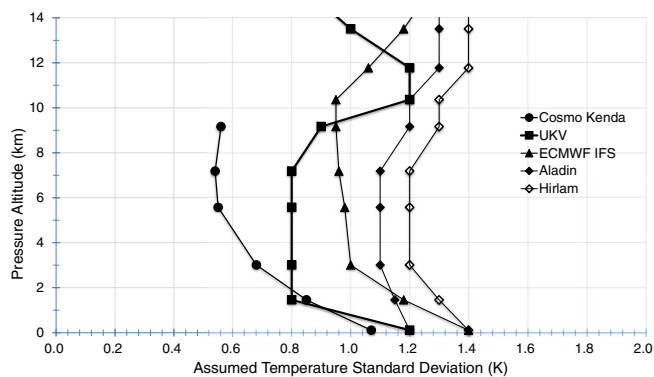
system for the United Kingdom (UKV) (Dalby and Berney, 1999, table 1, p. 4); European Centre for Medium Range Weather Forecasting integrated forecast system (IFS) (ECMWF, 2015, table 2.8, p. 44); Aire Limitée Adaptation dynamique Développement InterNational (ALADIN) (Strajnar, 2015); the High Resolution Limited Area Model (HIRLAM) (Unden et al., 2002, appendix B, p. 135).

2.2 | Empirical estimates

Empirical estimates are from two atmospheric observing systems that record measurements at a single point, for example, AMDAR and Radiosonde reports. Ideally, the two observing systems are in close proximity to each other in space and time to reduce the effects of mesoscale atmospheric variability. The strengths of this method are that it is independent of the NWP system and can be used to identify the range of useful observations for new instrumentation. The weaknesses are that it is limited by the total sample size, weather types and costs to acquire the data. Furthermore, this method accounts mostly only for the instrumental error and some residual small-scale mesoscale variability. For the empirical estimates we use, it is assumed that mesoscale variability is considered to have a small effect when the space–time separation between the observing systems is less than 40 km horizontally, 30 m vertically, and 15 min in time. Similarly for aircraft-based observations obtained using the method of flight-following, (e.g., in Stone (2018) the FAAM BAe-146 was used to follow a British Airways commercial aircraft over the Bristol Channel), the space–time separation along the same flight track the separation is around 10 km horizontally, 300 m vertically, and 60 s. For in situ data (collected from the flight data recorder for flights in and out of London Heathrow) provide observations at the same point in space and time. Likewise, Mirza et al. (2016) used a metrological method to compare in situ atmospheric measurements, made with research-grade instruments aboard the FAAM BAe-146, with (emulated) Mode-S EHS reports, at the same point in space and time. Figure 2 shows empirical estimates of observation uncertainty standard deviation for aircraft temperature reports from a range of studies: AMDAR versus Radiosonde (Schwartz and Benjamin, 1995; Ding et al., 2015, 2018); AMDAR versus other nearby AMDAR reporting aircraft (Benjamin et al., 1999); Mode-S MRAR versus Radiosonde (Strajnar, 2012), Mode-S EHS versus Radiosonde (de Haan and Stoffelen, 2012).

In broad terms, we summarize these empirical estimates for the temperature observation uncertainty as follows: for all sources except T_{MACH}, between 2 and 8 km the uncertainty is approximately between 0.75 and
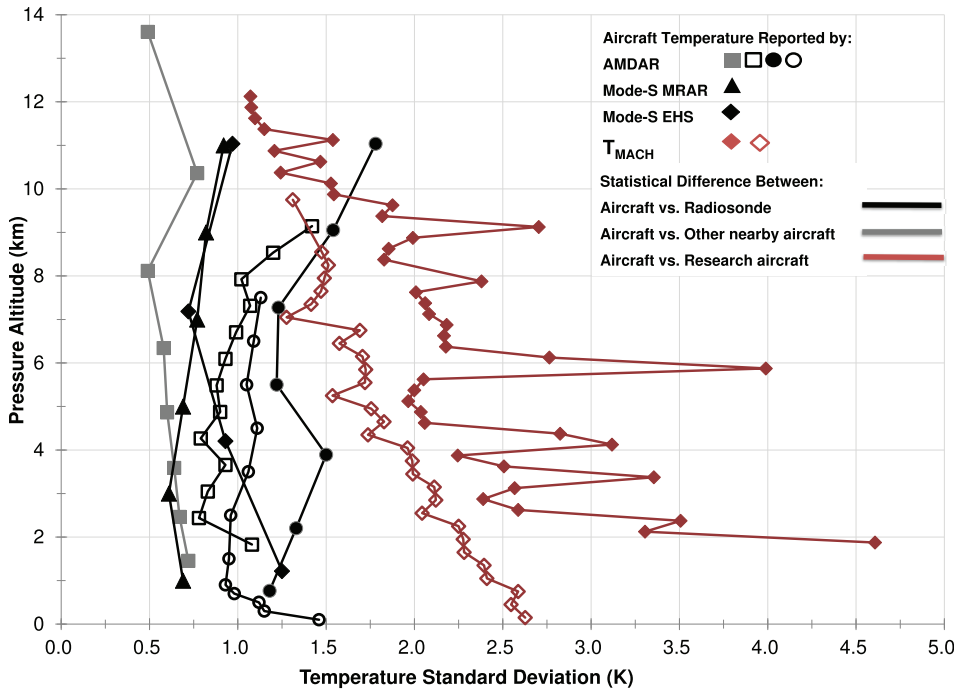

FIGURE 1 Vertical profile of assumed estimates of observation uncertainty for temperature reported by AMDAR aircraft, as used in various NWP data assimilation systems

FIGURE 2 Vertical profile of empirical estimates of observation uncertainty for temperature reported by aircraft. Symbols: AMDAR: Squares □ Schwartz and Benjamin (1995) ▪ Benjamin *et al.* (1999) and circles ● Ding *et al.* (2015) ○ Ding *et al.* (2018), Mode-S MRAR: Triangles ▲ Strajnar (2012), Mode-S EHS: Black diamond ◆ de Haan (2011), T_{MACH}: Red diamonds ◇ Mirza *et al.* (2016) ◆ Stone (2018)

1.25 K. It increases below 2 and above 8 km. For all altitudes, the uncertainty for T_{MACH} is approximately double that of the other sources, and can be characterized by a near linear decreasing trend with altitude.

Figure 2 also shows the results of the study by Mirza *et al.* (2016); here T_{MACH} were derived from emulated Mode-S EHS reports. Comparing these with in situ measured values made at the same time, the resulting empirical estimate shows a near linear decrease in observation error standard deviation with increasing altitude. Also shown is the study by Stone (2018), which used a similar method except using a commercial aircraft. The empirical estimate from this study also shows a decrease in observation error standard deviation with increasing altitude, albeit with slightly larger values and some outliers, especially at low altitude.

2.3 | Diagnosed estimates

Diagnosed estimates are made by comparing an atmospheric observing system and the output from an NWP data assimilation system, for example, AMDAR reports and the UKV NWP system. This method calculates statistics from samples of observation-minus-background (e.g., Hollingsworth and Lonnberg, 1986), $\mathbf{d}_{ob}$, called the innovation vector (Talagrand, 1997), and observation-minus-analysis, $\mathbf{d}_{oa}$, called the residual vector (e.g., Desroziers *et al.*, 2005), which must be acquired from the data assimilation process of an NWP system. Here we use the Desroziers *et al.* (2005)

diagnosis method to compute the statistical average between $\mathbf{d}_{ob}$ and $\mathbf{d}_{oa}$. This method results in an estimate of the statistics of the total error, combining the errors from the observation instruments and the NWP system.

2.3.1 | Desroziers *et al.*'s diagnosed estimate

We can define the innovation and residual for each observation as vectors of differences,

$$\mathbf{d}_{ob} = \mathbf{y}_o - \mathbf{H}(\mathbf{x}_b), \tag{1}$$

and

$$\mathbf{d}_{oa} = \mathbf{y}_o - \mathbf{H}(\mathbf{x}_a), \tag{2}$$

where $\mathbf{y}_o$ is the vector of observations, $\mathbf{H}$ is the observation operator that the data assimilation system uses to transform NWP model values into equivalent "real world" observations, and $\mathbf{x}_b$ and $\mathbf{x}_a$ are the corresponding vectors of model values before and after all observations have been assimilated, respectively.

Using linear statistical expectation theory (Walpole *et al.*, 2011, Ch. 4), and the assumption that the errors in NWP model values and the observations are mutually uncorrelated, Desroziers *et al.* (2005) showed that an estimate of the observation covariance error, $\mathbf{R}_e$, can be diagnosed from the statistical average, $E[]$, of the product $\mathbf{d}_{oa}(\mathbf{d}_{ob})^T$, that is,

$$E\left[\mathbf{d}_{oa}(\mathbf{d}_{ob})^T\right] \approx \mathbf{R}_e, \qquad (3)$$

where T is the matrix transpose operator and $\mathbf{R}_e$ are the covariances of the observation errors expressed as a square matrix. The elements along the diagonal of the square matrix are the observation error variances, σ_{ii}^2. We can compute these diagonal elements using,

$$\sigma_{ii}^2 = \frac{1}{N}\sum_{i=1}^{N}(d_{oa})_i(d_{ob})_i - \left(\frac{1}{N}\sum_{i=1}^{N}(d_{oa})_i\right)\left(\frac{1}{N}\sum_{i=1}^{N}(d_{ob})_i\right), \qquad (4)$$

where N is the total number of observations and i is the ith element along the diagonal of $\mathbf{R}_e$. To ensure the result of the diagnostic is unaffected by bias, the mean of the residual and innovation vectors are subtracted (Stewart, 2010; Waller et al., 2016b).

The diagnostic in Equation (3) only provides a correct estimate of the error covariance matrix if the assumed error statistics for the NWP model values and observations, used in the data assimilation processing, exactly represent the true statistics. Furthermore, the statistical construction of the diagnostic results in a non-symmetric matrix. Hence, any estimated error covariance matrix must be made symmetric before it can be used in an NWP system. There are further limitations of the diagnostic that can affect the estimated error covariance matrix, for example, the simplifying assumption of linear observation operators (Terasaki and Miyoshi, 2014); the use of localisation in data assimilation (Waller et al. 2017); and, in order to obtain sufficient sample residuals it is often assumed that uncertainties are ergodic, isotropic and homogeneous (Todling, 2015). Because of these (and possibly other) limitations of the diagnostic, error statistics estimated using this methodology should be interpreted as indicative, rather than necessarily quantitatively exact. Such results have nevertheless proved useful to identify the sources of observation and quality control errors (e.g., Waller et al., 2016a, 2016b, 2019).

With our definitions of these various methods used to estimate observation variances, in cases where empirical estimates are costly to obtain (so are not widely available nor updated frequently) the alternative methods of assumed and diagnosed estimates (which are easier to obtain) may instead be being used for the data assimilation of observations, for example, aircraft temperature reports. We suggest, however, that where empirical estimates are available they provide a reasonable basis to assess the validity of assumed and diagnosed estimates.

3 | EXPERIMENTAL METHODS

In this section, we give an overview of our experimental design to diagnose an estimate of the observation uncertainty for aircraft temperature reports obtained from AMDAR and Mode-S EHS. In Section 3.1 we outline the properties of the Met Office UKV NWP system and its three-dimensional variational data assimilation system (hereafter 3D-Var). The correction and selection of AMDAR and Mode-S EHS observations is described in Sections 3.2 and 3.3, and in Section 3.4, the 3D-Var used to obtain the innovation and residual vectors for use in the Desroziers et al. (2005) diagnostic method.

3.1 | UKV NWP system

The UKV NWP model is a limited-area, convection-permitting model configuration (Lean et al., 2008; Tang et al., 2013; Clark et al., 2016). At the time of this study the operational configuration was as follows:

- The NWP forecast model used a horizontal grid length of 1.5 km, 70 levels from the surface to the stratosphere.
- Routine observations from a range of surface, upper-air and satellite systems were assimilated (Ballard et al., 2016).
- A two-stage quality control was used: (a) removal of erroneous or invalid reports, (b) prior to assimilation removal of reports that are not consistent with other reports of the same type and reports whose probability of gross error is greater than 0.5 (Ingleby and Lorenc (1993) describes how this method is implemented for incremental 3D-VAR).
- Initial conditions were obtained from an incremental 3D-VAR scheme (Lorenc et al., 2000; Renshaw and Francis, 2011). This system uses a first guess at appropriate time (FGAT); a 3 km horizontal grid length; a latent heat nudging scheme (Jones and Macpherson, 1997); and boundary conditions obtained from the Global version of the Met Office Unified Model (Davies et al., 2005).
- An adaptive grid (Piccolo and Cullen, 2011) in the vertical was used, to resolve better boundary-layer features such as temperature inversions.
- The model ran eight times per day, forecasting up to 36 hr ahead.

The configuration of the UKV used for our experiment was the pre-operational parallel suite version 37 (UKV-PS37). For further details about the UKV model configuration and its 3D-Var scheme see Milan et al. (2019, section 3).

3.2 | Observation bias correction for aircraft temperature

The UKV routine pre-processing includes bias correction for AMDAR temperature reports for specific reporting aircraft (Ballish and Kumar, 2008). However, T_{MACH} reports are not bias-corrected since we found that the mean observation-minus-background was close to zero for the assimilated reports (Mirza, 2017, figure 7.10). Higher order corrections, such as those discussed by Zhu *et al.* (2015), could have been considered but were not available within the UKV data assimilation system used for this study.

3.3 | Observation thinning and selection

Previous studies have shown that the practical use of Mode-S EHS observations require thinning, removing duplicate or redundant observations, prior to their data assimilation (de Haan and Stoffelen, 2012; Lange and Janjic, 2016). Therefore, short-run trials were conducted to estimate the amount of data-thinning required. Our metric was simply the number of iterations required for the data assimilation to reach an acceptable level of convergence prior to the NWP forecasting step. In our experiments we used two methods: temporal thinning and spatial thinning (Mirza, 2017). For temporal thinning, the time-window for accepting observations is $T \pm \Delta t$ minutes, where T is the data assimilation time, for example, 0000, 0300, 0600 UTC, and Δt is the time window for accepting observations. The default time window for accepting observations is $\Delta t = 90$ min. The temporal thinning could be applied separately to the different types of aircraft-based observations. For our experiment, after trial and error, the time window for accepting T_{MACH} reports was set to $\Delta t = 30$ min. The time window for AMDAR was left at its default value, $\Delta t = 90$ min.

The spatial thinning grid-box dimensions were: 40 hPa vertical depth and the horizontal dimensions at the surface being 3.0 km in longitude and 3.0 km in latitude. The vertical depth of 40 hPa corresponds approximately to the vertical separation between aircraft flight levels (1,000 ft or 330 m) near the surface under International Standard Atmosphere conditions (ICAO, 1993). The time difference between reports within the grid box was 5 min. Spatial thinning is applied to all available aircraft-based observations. If there was more than one aircraft-based observation type available within a grid box then the T_{MACH} observation was preferentially selected, and after that the observation which is closest to the centre of a grid box was selected. The motivation for this choice arose because for the purposes of this study, our primary goal was to evaluate the uncertainty of the T_{MACH} data.

By these means the number of T_{MACH} reports was reduced to around 10% of those available, but this is still around 20 times more than the available AMDAR. This reduction is comparable to previous studies which used Mode-S EHS reports (de Haan and Stoffelen, 2012; Strajnar *et al.*, 2015; Lange and Janjic, 2016).

3.4 | Innovation and residuals from the UKV

For the UKV-PS37 data assimilation system, the assumed background error covariance matrices were the same as the existing operational suite (Ballard *et al.*, 2016, Figure 4). We used the UKV's assumed observation error variances for both AMDAR and T_{MACH} (see Figure 1). We assumed that these were the same because we aimed to obtain a configuration of the UKV-PS37 data assimilation system that would be similar to the operational suite. We also assumed that instrument errors for each reporting aircraft are uncorrelated with each other. Although this assumption may not be valid if there are aircraft-type specific biases (Drüe *et al.*, 2008) it is a reasonable first approximation. Furthermore, most operational data assimilation systems assume that spatial observations errors are uncorrelated. The data assimilation system used, unless thinned, all the available T_{MACH} reports derived from Mode-S EHS received at Met Office Mode-S EHS receiver sited at Thurnham (Stone and Pearce, 2016). All AMDAR reports received over the UKV domain were used. The reports used were collected between 0300 2 January to 0600 January 8, 2015 UTC, and were retrieved from the Met Office observations archive.

The UKV-PS37 was configured to initialize its run from the operational analysis obtained 3 hr earlier at 0000 UTC. This was to allow the NWP model a short spin-up time before starting to assimilate T_{MACH} reports. Moreover, the start time of the trial was chosen to allow for the gradual increase in the number of available T_{MACH} reports. This is because there is minimal air traffic operating within UK airspace between 2300 and 0500 UTC, other than aircraft transiting UK airspace at high altitude (≈ 10 km).

At the end of each data assimilation cycle the $\mathbf{d}_{ob}$ and $\mathbf{d}_{oa}$ for all observation types were stored for subsequent processing using the Desroziers *et al.* (2005) diagnostic method described in section 2.3.1.
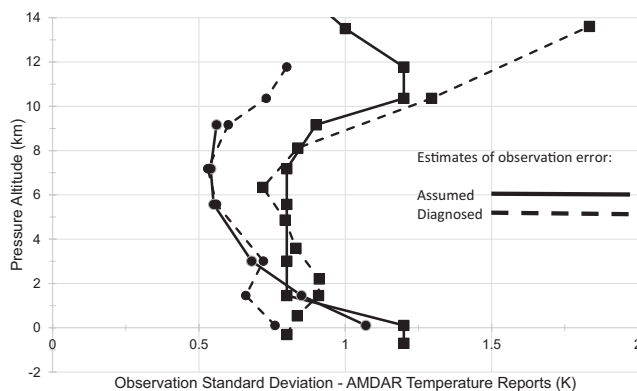
4 | RESULTS

We noted in Section 1 that the accurate representation of the observation uncertainty is important for the data

assimilation process. In this section we compare the vertical profiles of the Desroziers *et al.* (2005) diagnosed estimates of observation error variances for AMDAR temperature and T_{MACH} from the UKV data assimilation processing, shown in Figures 3 and 4, respectively, with the empirical estimates shown in Figure 2 and the assumed estimates shown in Figure 1. As noted in Section 3.2, the diagnosed estimates of observation bias for AMDAR temperature and T_{MACH} were near zero so, for clarity, these are not shown.

Figure 3 shows the vertical profiles for the assumed and diagnosed estimates of observation error variances for AMDAR temperature for the UKV. Firstly we see that the UKV diagnosed profile is in good agreement with the assumed profile between altitudes 2 and 10 km. The small differences are probably due to the relatively short

sampling period used (5 days). The increase in uncertainty above 10 km is probably due to the small sample of reports at these altitudes since there are fewer AMDAR reporting aircraft. The decrease in the observation uncertainty below 2 km is probably due to the time of year: the sampling period included periods of calm conditions resulting in temperature inversions so there would be little or no low-level turbulence. For comparison, Figure 3 also shows the assumed and diagnosed estimates for COSMO-KENDA (Lange and Janjic, 2016), for which similar conclusions may be drawn. Where the UKV and COSMO-KENDA differ may be due to their differences in their data assimilation processing, NWP model configurations and time of year (2–8 January 2015 and 7–12 May 2014, respectively). Nonetheless, these results serve to demonstrate how the Desroziers *et al.* (2005) diagnosis method can be used to check the consistency of observation uncertainty estimates for a particular data assimilation framework.

Figure 4 shows the vertical profiles for the assumed and diagnosed estimates of observation uncertainty for T_{MACH} for the UKV. It is clear that the diagnosed estimate is at least two times larger than the assumed estimate, which is used for AMDAR temperature reports. Furthermore, between the altitudes 2 and 10 km the diagnosed estimate shows a decreasing trend with increasing altitude. The increase in uncertainty above 10 km, again, may be due to fewer reporting aircraft. However, the increase in uncertainty below 2 km is most likely due to a combination of aircraft manoeuvres, flap configuration (Drüe *et al.*, 2008) and reduced precision of the reported Mach number (Mirza *et al.*, 2016). The shape of the observation uncertainty profile is unlikely to be affected by the assumed background error standard


FIGURE 3 Estimated UKV (■) observation error for AMDAR temperature reports: Assumed and diagnosed. For comparison, also shown are the corresponding results for COSMO-KENDA (●) (Lange and Janjic, 2016)

FIGURE 4 Estimated UKV (■) observation error for T_{MACH} reports: Assumed and diagnosed. For comparison also shown are, (a) from Figure 2, the empirical results for aircraft vs research aircraft (◇ ◆) and (b) corresponding results for COSMO-KENDA (●) smoothed Mode-S EHS reports (Lange and Janjic, 2016)
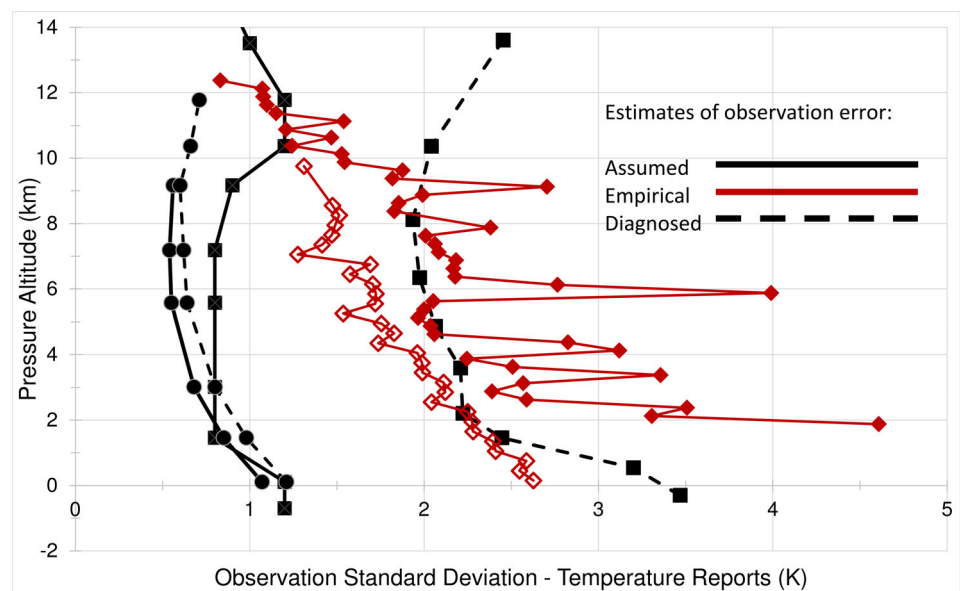
deviation since this does not vary greatly over the profile, being between 0.4 and 0.6 K (Ballard *et al.*, 2016, Figure 4). For comparison, Figure 4 also shows empirical estimates of the observation uncertainty from two different metrological studies. The UKV diagnosed estimate shows good correspondence with these metrological studies. We conclude therefore that the Desroziers *et al.* (2005) diagnosis method indicates that the assumed estimate of observation error variance for T_{MACH} used by the UKV is inconsistent with its diagnosed estimate. This result is only indicative, we suggest a longer study is needed to validate the result.

Figure 4 also shows the assumed and diagnosed estimates of observation error variance of Mode-S EHS temperatures for COSMO-KENDA (Lange and Janjic, 2016). Here the Desroziers *et al.* (2005) diagnosed estimate is consistent with the assumed estimate. Both these estimates are also in close agreement with the UKV assumed estimate for AMDAR temperature reports. A key difference between UKV and the COSMO-KENDA diagnosed estimates of observation error variance is likely to be due to the difference in the data used: the T_{MACH} reports for the UKV used unsmoothed Mode-S EHS reports whereas COSMO-KENDA used reports that had been smoothed using a low-pass filter based on linear regression (de Haan, 2011; Mirza *et al.*, 2019); other key differences of COSMO-KENDA are its horizontal grid length of 8 km, 30 vertical levels and a 40-member ensemble-based Kalman filter to sample the NWP model background error covariance (see Lange and Janjic, 2016, Section 3).

We conclude from these results that the close agreement with the metrological studies gives confidence in the quantitative values of the diagnosed standard deviations in this study.

5 | SUMMARY

In this paper we used the Desroziers *et al.* (2005) diagnosis method to compute new diagnosed estimates of the vertical structure of the uncertainty variances for AMDAR and T_{MACH} temperatures, using the observation-minus-background and observation-minus-analysis values output from the Met Office UKV 3D-VAR. We compared the consistency of these diagnosed estimates with those obtained from previous studies.

We did this firstly by comparing assumed estimates with corresponding empirical estimates of observation uncertainty for AMDAR temperature reports. We then used the Desroziers *et al.* (2005) diagnosis method to obtain a diagnosed estimate of the AMDAR temperature observation uncertainty for the UKV. We showed that the assumed and diagnosed estimates are consistent with each

other. Secondly, we applied the Desroziers *et al.* (2005) diagnosis method to T_{MACH} reports, derived from Mode-S EHS, to obtain a new diagnosed estimate of their observation uncertainty for the UKV. We showed that, in this case, the assumed and diagnosed estimates were inconsistent, the newly diagnosed estimate being two times larger than the assumed estimate. Moreover, we showed that the newly diagnosed estimate is consistent with corresponding empirical estimates.

We noted in our introduction that there are very few studies that have compared Desroziers *et al.* (2005)-type observation uncertainty estimates with independent estimates of the observation uncertainty. We have shown that the Desroziers *et al.* (2005) diagnosis method is a valuable tool for checking the consistency of the assumed observation uncertainties and for estimating the observation standard deviation for use in an NWP system. Mirza *et al.* (2019) showed that the variance of Mach temperatures is reduced when a smoothing filter is used, so that a reduction in the COSMO-KENDA values (Lange and Janjic, 2016) relative to those in the UKV is appropriate, given the different treatments of the data. However, as has been shown, the magnitudes of the assumed errors may be incorrect. Our analysis has been based on a short case study. Hence, if T_{MACH} reports are to be assimilated in the UKV NWP system then the corresponding assumed estimate of their observation uncertainty should be investigated further.

ORCID

Andrew K. Mirza https://orcid.org/0000-0001-6350-9080
Sarah L. Dance https://orcid.org/0000-0003-1690-3338
Gabriel G. Rooney https://orcid.org/0000-0002-3787-1198
David Simonin https://orcid.org/0000-0002-2623-3828
Edmund K. Stone https://orcid.org/0000-0002-0298-1979
Joanne A. Waller https://orcid.org/0000-0002-7783-6434

REFERENCES

Ballard, S.P., Zhihong, L., Simonin, D. and Caron, J.-F. (2016) Performance of 4D-Var NWP-based nowcasting of precipitation at the Met Office for summer 2012. *Quarterly Journal of the Royal Meteorological Society*, 142(694), 472–487.

Ballish, B.A. and Kumar, V.K. (2008) Systematic differences in aircraft and radiosonde temperatures implications for NWP and climate studies. *Bulletin of the American Meteorological Society*, 89(11), 1689–1708. https://doi.org/10.1175/2008BAMS 2332.1.

Bathmann, K. (2018) Justification for estimating observation-error covariances with the Desroziers diagnostic. *Quarterly Journal of the Royal Meteorological Society*, 144(715), 1965–1974.

Benjamin, S.G., Schwartz, B.E. and Cole, R.E. (1999) Accuracy of ACARS wind and temperature observations determined by collocation. *Weather and Forecasting*, 14(6), 1032–1038. https://doi.org/10.1175/1520-0434(1999)014<1032:AOAWAT>2.0.CO;2.

Bormann, N., Bonavita, M., Dragani, R., Eresmaa, R., Matricardi, M. and McNally, A. (2016) Enhancing the impact of IASI observations through an updated observation-error covariance matrix. *Quarterly Journal of the Royal Meteorological Society*, 142(697), 1767–1780. https://doi.org/10.1002/qj.2774.

Campbell, W.F., Satterfield, E.A., Ruston, B. and Baker, N.L. (2017) Accounting for correlated observation error in a dual-formulation 4D variational data assimilation system. *Monthly Weather Review*, 145(3), 1019–1032. https://doi.org/10.1175/MWR-D-16-0240.1.

Cardinali, C., Isaksen, L. and Andersson, E. (2003) Use and impact of automated aircraft data in a global 4DVAR data assimilation system. *Monthly Weather Review*, 131(8), 1865–1877. https://doi.org/10.1175//2569.1.

Chun, H.-W., Eresmaa, R., McNally, A. P., Bormann, N. and Matricardi, M. (2015), A physically-based observation error covariance matrix for IASI, *in* 'The 20th International TOVS Study Conference (ITSC-20) Lake Geneva, Wisconsin, USA https://cimss.ssec.wisc.edu/itwg/itsc/itsc20/program/

Clark, P., Roberts, N., Lean, H., Ballard, S.P. and Charlton-Perez, C. (2016) Convection-permitting models: a step-change in rainfall forecasting. *Meteorological Applications*, 23(2), 165–181.

Dalby, T. D. and Berney, A. (1999), Modifications to aircraft thinning and observation errors., Forecasting Research Technical Report 276, Met Office, Exeter, United Kingdom. https://digital.nmla.metoffice.gov.uk/IO_41e315e67661446a929a2db64 f9db8b5/

Davies, T., Cullen, M.J.P., Malcolm, A.J., Mawson, M.H., Staniforth, A., White, A.A. and Wood, N. (2005) A new dynamical core for the Met Office's global and regional modelling of the atmosphere. *Quarterly Journal of the Royal Meteorological Society*, 131(608), 1759–1782. https://rmets.onlinelibrary.wiley.com/doi/abs/10.1256/qj.04.101.

de Haan, S. (2011) High-resolution wind and temperature observations from aircraft tracked by Mode-S air traffic control radar. *Journal of Geophysical Research*, 116(D10). https://doi.org/10.1029/2010jd015264.

de Haan, S. and Stoffelen, A. (2012) Assimilation of high-resolution Mode-S EHS wind and temperature observations in a regional NWP model for nowcasting applications. *Weather and Forecasting*, 27(4), 918–937. https://dx.doi.org/10.1175/WAF-D-11-00088.1.

Desroziers, G., Berre, L., Chapnik, B. and Poli, P. (2005) Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society*, 131(613), 3385–3396. https://doi.org/10.1256/qj.05.108.

Ding, J., Zhuge, X., Li, X., Yuan, Z. and Wang, Y. (2018) Evaluation of accuracy of chinese AMDAR data for 2015. *Journal of Atmospheric and Oceanic Technology*, 35(5), 943–951. https://doi.org/10.1175/JTECH-D-17-0095.1.

Ding, J., Zhuge, X., Wang, Y. and Xiong, A. (2015) Evaluation of chinese aircraft meteorological data relay (AMDAR) weather reports. *Journal of Atmospheric and Oceanic Technology*, 32(5), 982–992. https://doi.org/10.1175/JTECH-D-14-00145.1.

Drüe, C., Frey, W., Hoff, A. and Hauf, T. (2008) Aircraft type-specific errors in AMDAR weather reports from commercial aircraft. *Quarterly Journal of the Royal Meteorological Society*, 134(630), 229–239.

ECMWF (2015), IFS Documentation cy41r1 operational implementation May 12, 2015 part i: observations, Technical report, European Centre for Medium Range Weather Forecasting, Reading, United Kingdom. https://www.ecmwf.int/sites/default/files/elibrary/2015/9208-part-i-observation-processing.pdf

Hollingsworth, A. and Lonnberg, P. (1986) The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: the wind field. *Tellus A*, 38A(2), 111–136. https://doi.org/10.1111/j.1600-0870.1986.tb00460.x.

ICAO. (1993) Manual of the ICAO standard atmosphere: extended to 80 Kilometres (262 500 feet). In: *Third Edition, Technical Report Doc 7488-CD*. Montreal, Canada: International Civil Aviation Organisation.

Ingleby, B., Isaksen, L. and Kral, T. (2019), 'Evaluation and impact of aircraft humidity data in ECMWF's NWP system', *ECMWF Technical Memoranda* (855) https://dx.doi.org/10.21957/4e825dtiy

Ingleby, N.B. and Lorenc, A.C. (1993) Bayesian quality control using multivariate normal distributions. *Quarterly Journal of the Royal Meteorological Society*, 119(513), 1195–1225. https://doi.org/10.1002/qj.49711951316.

James, E.P. and Benjamin, S.G. (2017) Observation system experiments with the hourly updating rapid refresh model using GSI hybrid ensemble-variational data assimilation. *Monthly Weather Review*, 145(8), 2897–2918. https://doi.org/10.1175/MWR-D-16-0398.1.

Janjic, T., Bormann, N., Bocquet, M., Carton, J.A., Cohn, S.E., Dance, S.L., Losa, S.N., Nichols, N.K., Potthast, R., Waller, J.A. and Weston, P. (2018) On the representation error in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 144(713), 1257–1278.

Jones, C.D. and Macpherson, B. (1997) A latent heat nudging scheme for the assimilation of precipitation data into an operational mesoscale model. *Meteorological Applications*, 4(3), 269–277. https://doi.org/10.1017/S1350482797000522.

Lange, H. and Janjic, T. (2016) Assimilation of Mode-S EHS aircraft observations in COSMO-KENDA. *Monthly Weather Review*, 144 (5), 1697–1711. https://doi.org/10.1175/MWR-D-15-0112.1.

Lean, H.W., Clark, P.A., Dixon, M., Roberts, N.M., Fitch, A., Forbes, R. and Halliwell, C. (2008) Characteristics of high-resolution versions of the Met Office unified model for forecasting convection over the United Kingdom. *Monthly*

Weather Review, 136(9), 3408–3424. https://doi.org/10.1175/2008MWR2332.1.

Lorenc, A.C., Ballard, S.P., Bell, R.S., Ingleby, N.B., Andrews, P.L. F., Barker, D.M., Bray, J.R., Clayton, A.M., Dalby, T., Li, D., Payne, T.J. and Saunders, F.W. (2000) The Met Office global three-dimensional variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, 126(570), 2991–3012. https://doi.org/10.1002/qj.49712657002.

Lorenc, A.C. and Marriott, R.T. (2014) Forecast sensitivity to observations in the Met Office global numerical weather prediction system. *Quarterly Journal of the Royal Meteorological Society*, 140(678), 209–224. https://doi.org/10.1002/qj.2122.

Ménard, R. (2016) Error covariance estimation methods based on analysis residuals: theoretical foundation and convergence properties derived from simplified observation networks. *Quarterly Journal of the Royal Meteorological Society*, 142, 257–273. https://doi.org/10.1002/qj.2650.

Milan, M., Macpherson, B., Tubbs, R., Dow, G., Inverarity, G., Mittermaier, M., Halloran, G., Kelly, G., Li, D., Maycock, A., Payne, T., Piccolo, C., Stewart, L. and Wlasak, M. (2019) Hourly 4D-var in the Met Office UKV operational forecast model. *Quarterly Journal of the Royal Meteorological Society*. 1281–1301. https://doi.org/10.1002/qj.3737.

Mirza, A. K. (2017), *On the utilization of aircraft derived observations for operational meteorology and numerical weather prediction*. PhD thesis, Department of Meteorology, University of Reading, United Kingdom. http://centaur.reading.ac.uk/71981/

Mirza, A.K., Ballard, S.P., Dance, S.L., Maisey, P., Rooney, G.G. and Stone, E.K. (2016) Comparison of aircraft-derived observations with in situ research aircraft measurements. *Quarterly Journal of the Royal Meteorological Society*, 142(701), 2949–2967. https://doi.org/10.1002/qj.2864.

Mirza, A.K., Ballard, S.P., Dance, S.L., Rooney, G.G. and Stone, E. K. (2019) Towards operational use of aircraft-derived observations: a case study at London Heathrow airport. *Meteorological Applications*, 26, 542–555.

Petersen, R.A. (2016) On the impact and benefits of AMDAR observations in operational forecasting. Part I: a review of the impact of automated aircraft wind and temperature reports. *Bulletin of the American Meteorological Society*, 97(4), 585–602. https://doi.org/10.1175/BAMS-D-14-00055.1.

Piccolo, C. and Cullen, M. (2011) Adaptive mesh method in the Met Office variational data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656), 631–640. https://doi.org/10.1002/qj.801.

Renshaw, R. and Francis, P.N. (2011) Variational assimilation of cloud fraction in the operational Met Office unified model. *Quarterly Journal of the Royal Meteorological Society*, 137(661), 1963–1974. https://doi.org/10.1002/qj.980.

Schraff, C., Reich, H., Rhodin, A., Schomburg, A., Stephan, K., Periáñez, A. and Potthast, R. (2016) Kilometre-scale ensemble data assimilation for the COSMO model (KENDA). *Quarterly Journal of the Royal Meteorological Society*, 142(696), 1453–1472.

Schwartz, B. and Benjamin, S.G. (1995) A comparison of temperature and wind measurements from ACARS-equipped aircraft and rawinsondes. *Weather and Forecasting*, 10(3), 528–544.

Simonin, D., Waller, J.A., Ballard, S.P., Dance, S.L. and Nichols, N. K. (2019) A pragmatic strategy for implementing spatially

correlated observation errors in an operational system: an application to Doppler radial winds. *Quarterly Journal of the Royal Meteorological Society*, 145(723), 2772–2790.

Stewart, L. M. (2010), *Correlated observation errors in data assimilation*. PhD thesis, Departments of Mathematics and Meteorology, University of Reading. http://www.reading.ac.uk/web/files/maths/StewartPhD2010.pdf [accessed 11 February 2021].

Stewart, L.M., Dance, S.L. and Nichols, N.K. (2008) Correlated observation errors in data assimilation. *International Journal for Numerical Methods in Fluids*, 56(8), 1521–1527. https://doi.org/10.1002/fld.1636.

Stewart, L.M., Dance, S.L. and Nichols, N.K. (2013) Data assimilation with correlated observation errors: experiments with a 1-D shallow water model. *Tellus A*, 65. https://doi.org/10.3402/tellusa.v65i0.19546.

Stewart, L.M., Dance, S.L., Nichols, N.K., Eyre, J.R. and Cameron, J. (2014) Estimating interchannel observation-error correlations for IASI radiance data in the Met Office system. *Quarterly Journal of the Royal Meteorological Society*, 140(681), 1236–1244. https://doi.org/10.1002/qj.2211.

Stickland, J. and Grooters, A. T. F. (2005), The global AMDAR programme, *in* 'International Symposium on Remote Sensing of the Environment'. http://www.isprs.org/proceedings/2005/ISRSE/html/papers/404.pdf

Stone, E.K. (2018) A comparison of Mode-S Enhanced Surveillance observations with other in situ aircraft observations. *Quarterly Journal of the Royal Meteorological Society*, 144(712), 695–700. https://doi.org/10.1002/qj.3238.

Stone, E.K. and Kitchen, M. (2015) Introducing an approach for extracting temperature from aircraft GNSS and pressure altitude reports in ADS-B messages. *Journal of Atmospheric and Oceanic Technology*, 32(4), 736–743. https://doi.org/10.1175/JTECH-D-14-00192.1.

Stone, E.K. and Pearce, G. (2016) A Network of Mode-S Receivers for Routine Acquisition of Aircraft-Derived Meteorological Data. *Journal of Atmospheric and Oceanic Technology*, 33(4), 757–768. https://doi.org/10.1175/jtech-d-15-0184.1.

Strajnar, B. (2012), Validation of Mode-S meteorological routine air report aircraft observations. *J. Geophys. Res*, 117, D23110, https://doi.org/10.1029/2012JD018315.

Strajnar, B. (2015), *Four-dimensional data assimilation of aircraft observations of the atmosphere in complex terrain*. PhD thesis, Department of Physics, University of Ljubljana. https://doi.org/10.13140/RG.2.1.3894.4081.

Strajnar, B., Aagar, N. and Berre, L. (2015) Impact of new aircraft observations Mode-S MRAR in a mesoscale NWP model. *Journal of Geophysical Research-Atmospheres*, 120(9), 3920–3938. https://doi.org/10.1002/2014JD022654.

Talagrand, O. (1997) Assimilation of observations, an introduction. *Journal of the Meteorological Society of Japan. Ser. II*, 75(1B), 191–209. https://www.jstage.jst.go.jp/article/jmsj1965/75/1B/75_1B_191/_article.

Tang, Y., Lean, H. and Bornemann, J. (2013) The benefits of the Met Office variable resolution NWP model for forecasting convection. *Meteorological Applications*, 20(4), 417–426. https://doi.org/10.1002/met.1300.

Tavolato, C. and Isaksen, L. (2015) On the use of a Huber norm for observation quality control in the ECMWF 4D-Var. *Quarterly*

Journal of the Royal Meteorological Society, 141(690), 1514–1527. https://doi.org/10.1002/qj.2440.

Terasaki, K. and Miyoshi, T. (2014) Data assimilation with error-correlated and non-orthogonal observations: experiments with the Lorenz-96 model. *Scientific Online Letters on the Atmosphere*, 10, 210–213. https://doi.org/10.2151/sola.2014-044.

Todling, R. (2015) A complementary note to 'A lag-1 smoother approach to system-error estimation': the intrinsic limitations of residual diagnostics. *Quarterly Journal of the Royal Meteorological Society*, 141, 2917–2922. https://doi.org/10.1002/qj.2546.

Unden, P., Rontu, L., Jarvinen, H., Lynch, P., Calvo, J., Cats, G., Cuxart, J., Eerola, K., Fortelius, C., Garcia-Moya, J. A., Jones, C., Lenderlink, G., Mcdonald, A., Mcgrath, R., Navascues, B., Nielsen, N. W., Degaard, V., Rodriguez, E., Rummukainen, M., Sattler, K., Sass, B. H., Savijarvi, H., Schreur, B. W., Sigg, R. and The, H. (2002), HIRLAM-5 Scientific documentation, Technical report. http://hirlam.org/index.php/hirlam-programme-53/general-model-description/synoptic-scale-hirlam

Waller, J.A., Ballard, S.P., Dance, S.L., Nichols, N.K., Simonin, D. and Kelly, G. (2016b) Diagnosing horizontal and inter-channel observation error correlations for SEVIRI observations using observation-minus-background and observation-minus-analysis statistics. *Remote Sensing*, 8(7), 581. http://www.mdpi.com/2072-4292/8/7/581.

Waller, J.A., Bauernschubert, E., Dance, S.L., Nichols, N.K., Potthast, R. and Simonin, D. (2019) Observation error statistics for Doppler radar radial wind superobservations assimilated into the DWD COSMO-KENDA system. *Monthly Weather Review*, 147(9), 3351–3364. https://doi.org/10.1175/MWR-D-19-0104.1.

Waller, J.A., Dance, S.L. and Nichols, N.K. (2016) Theoretical insight into diagnosing observation error correlations using observation-minus-background and observation-minus-analysis statistics. *Quarterly Journal of the Royal Meteorological Society*, 142(694), 418–431. https://doi.org/10.1002/qj.2661.

Waller, J.A., Dance, S.L. and Nichols, N.K. (2017) On diagnosing observation error statistics in localized ensemble data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 143 (708), 2677–2686. https://doi.org/10.1002/qj.3117.

Waller, J.A., Simonin, D., Dance, S.L., Nichols, N.K. and Ballard, S. P. (2016a) Diagnosing observation error correlations for Doppler radar radial winds in the Met Office UKV model using observation-minus-background and observation-minus-analysis statistics. *Monthly Weather Review*, 144(10), 3533–3551. https://doi.org/10.1175/MWR-D-15-0340.1.

Walpole, R.E., Myers, R.H., Myers, S.L. and Ye, K. (2011) *Probability & Statistics for Engineers & Scientists*, 9th edition. Boston, MA: Pearson Education.

Weston, P.P., Bell, W. and Eyre, J.R. (2014) Accounting for correlated error in the assimilation of high-resolution sounder data. *Quarterly Journal of the Royal Meteorological Society*, 140(685), 2420–2429. https://doi.org/10.1002/qj.2306.

WMO. (2017) *Guide to Aircraft-Based Observations, WMO-No 1200. 2017 edition*. World Meteorological Organisation, Geneva: Switzerland. https://library.wmo.int/doc_num.php?explnum_id=4120.

Zhu, Y., Derber, J.C., Purser, R.J., Ballish, B.A. and Whiting, J. (2015) Variational correction of aircraft temperature bias in the NCEPs GSI analysis system. *Monthly Weather Review*, 143(9), 3774–3803. https://doi.org/10.1175/MWR-D-14-00235.1.