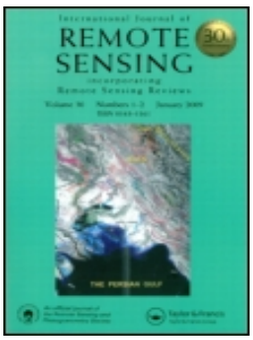

International Journal of Remote Sensing

ISSN: 0143-1161 (Print) 1366-5901 (Online) Journal homepage: http://www.tandfonline.com/loi/tres20

An operational tool to quality control 2D radar reflectivity data for assimilation in COSMO-DE

Kathleen Helmert , Birgit Hassler & Jörg E. E. Seltmann

To cite this article: Kathleen Helmert , Birgit Hassler & Jörg E. E. Seltmann (2012) An operational tool to quality control 2D radar reflectivity data for assimilation in COSMO-DE, International Journal of Remote Sensing, 33:11, 3456-3471, DOI: 10.1080/01431161.2011.592161

To link to this article: http://dx.doi.org/10.1080/01431161.2011.592161

Published online: 24 Nov 2011.

International Journal of Remote Sensing
Vol. 33, No. 11, 10 June 2012, 3456–3471

Taylor & Francis
Taylor & Francis Group

An operational tool to quality control 2D radar reflectivity data for assimilation in COSMO-DE

KATHLEEN HELMERT*†, BIRGIT HASSLER‡§¶
and JÖRG E. E. SELTMANN‡
†German Weather Service, Offenbach, Germany
‡German Weather Service, Meteorological Observatory Hohenpeissenberg, Hohenpeissenberg, Germany
§Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, USA
¶Chemical Sciences Division, NOAA Earth System Research Laboratory, Boulder, CO, USA

(*Received 7 May 2008; in final form 24 May 2011*)

An operational tool has been designed to enhance the quality of 2D radar reflectivity data for assimilation in COSMO-DE within the German Weather Service (DWD). This article describes the operational algorithms including their testing, the creation of local and composite quality-index fields and their application to improve data assimilation. In the first step, algorithms have been developed and tested to define and identify some of the most severe errors in radar data: corrupt images, low-reflectivity phenomena, which occur under special meteorological conditions, spokes, rings and clutter remnants/speckles. The algorithms use simple but effective tests based on statistical and textural characteristics of spurious signals. The results are stored in an 8-bit quality-index field created concurrently with the radar data at each radar station. It contains quality flags for each individual range bin plus some header information on the overall quality of the underlying data set. The independent coding of error bits enables a differentiated *a posteriori* decision on whether a range bin is to be used for a given application. These local quality index fields are then used to create a radar precipitation composite, accompanied by a quality-index composite, covering Germany. This tool has now been applied operationally throughout the German radar network. As a result, the radar data quality in data assimilation could be increased.

1. Introduction

The German Weather Service (DWD) operates a kilometre-scale numerical weather prediction system (COSMO-DE) for very short range forecasts (2–18 hours) (Baldauf *et al*. (2007) and references therein). COSMO-DE has a grid-spacing of 2.8 km with about 50 vertical layers and an integration domain of about 1300 × 1300 km, covering Germany and part of its neighbours. The system focuses on a model-based prediction of severe weather events on the meso-gamma scale, specifically those related to deep moist convection or to interactions with fine-scale topography. To trigger model deep convection, assimilation of meso-gamma structures is necessary. In COSMO-DE, the method of latent heat nudging is used (Jones and MacPherson 1997; Klink

*Corresponding author. Email: Kathleen.Helmert@dwd.de

International Journal of Remote Sensing
ISSN 0143-1161 print/ISSN 1366-5901 online © 2011 Taylor & Francis
http://www.tandf.co.uk/journals
http://dx.doi.org/10.1080/01431161.2011.592161

and Stephan 2004). For this reason, high-resolution radar precipitation composites are required and their quality control becomes crucial (Sun and Wilson 2003, Leuenberger 2005, Stephan *et al*. 2008). If corrupt radar data were admitted to assimilation, erroneous precipitation structures would propagate into the model fields (e.g. to the integrated water vapour field) for the next 2–3 hours. Quantitative accuracy is only second in importance compared to artefacts within the precipitation field (Leuenberger 2005, Stephan *et al*. 2008), because latent heat nudging is not considered as a quantitatively very accurate assimilation method, and, therefore, the demands on the accuracy of the observations are also limited. It is, however, important to remove artefacts in the observations that are typically related to gross errors and indeed can lead to large errors in the analysis.

Quality-controlled, high-resolution radar precipitation data are also vital for many other applications, such as flood forecasts or severe weather warning systems (Lang 2001, Laborda *et al*. 2006, Marx *et al*. 2006). Until recently, the quality of operationally available data sets did not always meet the user's predefined standards. Therefore, a majority of users built their own specific offline quality tests for radar data. Golz *et al*. (2006) gave a literature review of radar data quality control methods under the VOLTAIRE project sponsored by the European Community. Many of the methods are dependent on a variety of additional meteorological information and turn out to be time-consuming even in case studies. However, for the time-critical data assimilation in COSMO-DE, quality-controlled radar precipitation data should be available in near real time every 5 min. For real-time purposes, fast automated quality control procedures need to be applied.

Several of such methods for operational use have been described in the literature. Joss and Lee (1995) suggested an elaborate procedure for operational radar data processing in Switzerland. Harrison *et al.* (2000) presented quality control methods for the radar network at the UK Met Office as part of the Nimrod system. Peura (2002) and Michelson and Sunhede (2004) reported on data quality control in operational Scandinavian radar networks. However, most of them use radar data with a time resolution of more than 5 min and provide no information as to which range bins were contaminated and by what type of error. The method suggested by Friedrich *et al.* (2006) is the only one to deal with this issue, to the best of our knowledge. In their method, an averaged quality-index field is calculated for reflectivity, polarimetric parameters and Doppler velocity. Unfortunately, it has not yet become operational.

2. Concept and scope

Given the specific requirements of the COSMO-DE assimilation, and keeping in mind the different, at times contradicting aspects and quality standards of the many users asking for DWD weather radar data, the aim of this article is to provide a robust, flexible and developable compromise to operational quality assurance. To this end, the occurrence of major problems, as defined by radar users, should be detected and flagged without prejudicing a scale to their severity (quality measure), without mixing them up or even changing the data themselves. The idea is to operationally generate a quality-index field in which several detected errors for each range bin may be binary coded. This approach offers the possibility of combining different bits with different weights to form individual quality measures adapted to the user's need, who may wish to apply different correction and application algorithms, possibly automatically controlled by the errors detected. For instance, clear air echoes, once detected, might

be excluded from precipitation assimilation, but admitted to radial wind processing. Thus, the weight that may be given to each of the error types, resulting in a quality measure, is left to the user's discretion. This article outlines the error types considered and the corresponding detection algorithms used. It includes test results and the layout of the quality-index field, as well as an example of a quality measure in the above sense used to create a precipitation composite for COSMO-DE.

Severe radar data errors as judged by DWD radar users include corrupt images, spokes caused by beam blockage or external transmitters, anomalous propagation (anaprop) echoes, remnants of ground clutter and returns from ships or wind parks, which will all be explained in §4. The quality control procedure to handle these problems described in §4 is based on simple, robust and fast (capable of being automated easily), but nevertheless effective detection algorithms. At this stage, algorithms are mainly statistical in nature, or texture based; no sophisticated, physically based or quantitative correction is intended here. A later replacement or addition of modules, for instance algorithms for bright band or attenuation detection, is straightforward.

All identified errors for each range bin are stored in a quality-index field, which has the same resolution as the radar data field itself. There are different bit values reserved for different problems, which allow an individual bit combination for various applications. The generation of the quality index is run in a post-processing mode on the local radar host computers and is then distributed to individual users. Several evaluation tests have been performed which are described in §4.3. Furthermore, the quality-index fields of the individual radars are used to create a quality-controlled radar precipitation composite for data assimilation as described in §5.

3. Database

DWD operates a homogeneous network of 17 Doppler radars at the C-band (5640 MHz). Several types of commercial signal processors are used to process the raw radar data. Doppler filters are run on the time series data to quantitatively correct ground clutter (Seltmann 2000). Following integration, the data are quality checked using certain logic combinations of thresholds for power above noise, strength of clutter correction, weather signal power and signal coherency (Seltmann and Riedl 1998). Cluttered 125 m raw bins are excluded from range averaging (clutter micro suppression), and stand-alone range-averaged 1 km bins are discarded (radial speckle remover). Thresholded range bins are set to zero (which is defined as 'no data') by the signal processor.

Subsequently, three different scan modes are performed nearly synchronously by each individual radar. Volume data are available from both a Doppler scan (18 elevations at a staggered pulse repetition rate of 800/1200 Hz, azimuth rate = 3 rpm, range = 128 km) and an intensity scan (5 elevations at 600 Hz, 3 rpm, 256 km) every 15 min. A dedicated precipitation scan (one revolution at the lowest elevation angle possible following the horizon line, 600 Hz, 2 rpm, 125 km) is performed every 5 min with priority over Doppler and intensity scans. From this precipitation scan, an 8-bit product in polar coordinates (range r, azimuth ϕ) is obtained with a spatial resolution of $\Delta r = 1$ km, $\Delta\phi = 1^\circ$. These 2D products are the basis for all precipitation calculations near the surface and for the current work. The volume data have so far been exploited in neither hydrological nor model assimilation applications.

4. Identification of spurious precipitation signals

In spite of the filters described above, there may still be clutter remnants. Specifically, moving targets such as ships, aircraft or wind parks cannot be detected by Doppler filters. In addition, spokes and rings or ring segments, caused by obstacles or external transmitters, sometimes occur in the data. Problems caused by the radar itself (corrupt image (CI) or by special meteorological conditions (German pancake (GP): Hassler *et al*. (2005); or anaprop) tend to cause problems in data assimilation as well.

Therefore, to control radar data quality, a procedure was developed containing algorithms to identify gross errors and errors affecting the occurrence of precipitation as opposed to its quantification (Hassler *et al*. 2006). Even though variations of the vertical reflectivity profile, beam shielding and attenuation do strongly affect the estimation of the precipitation amount, as described by Joss and Waldvogel (1990), Doviak and Zrnic (1993) and Fabry *et al.* (1992), the quantitative accuracy of real precipitation is less crucial for the purpose of data assimilation than excluding spurious features (Leuenberger 2005, Stephan *et al*. 2008).

With our approach, the detection of radar errors is organized in different modules, which can be switched on and off separately. First, data are searched for problems that affect the entire data set, and then algorithms deal with errors that can be assigned to individual range bins. Figure 1 shows a schematic of the detection methods described in the following sections.

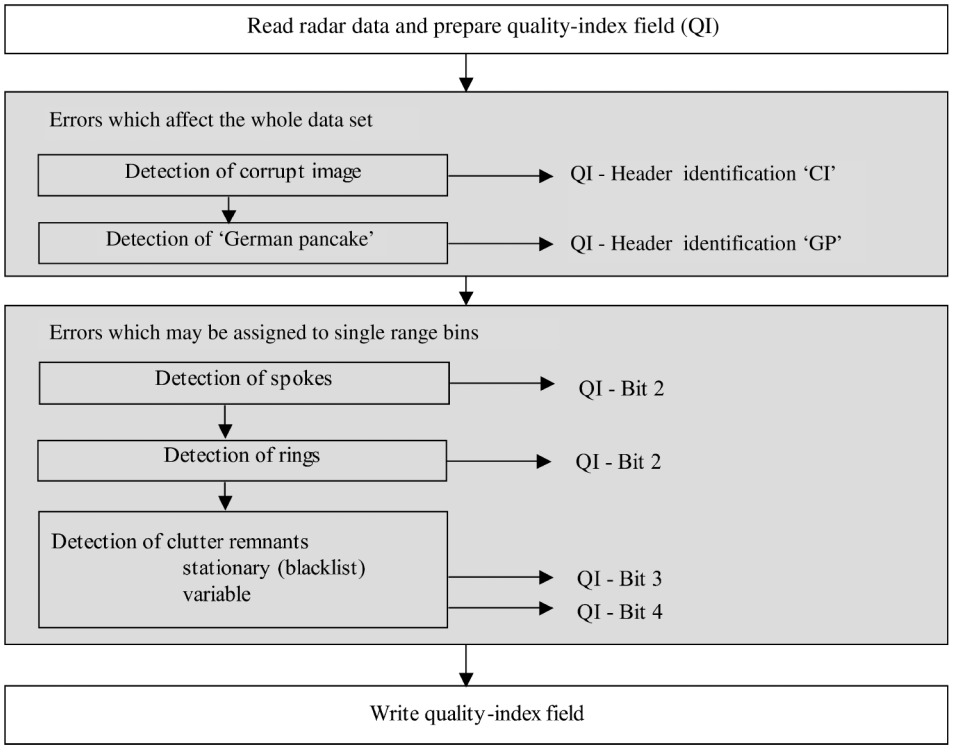

Figure 1. Flow chart with the presently available detection algorithms and the corresponding bits in the quality-index field.

4.1 *General problems*

Errors caused by the radar itself or by special meteorological conditions often affect the data set as a whole, or larger areas in it. It is impractical to assign these errors to single range bins. In this case, the identified errors are marked in the header of the quality-index field with the appropriate identifiers (CI or GP).

4.1.1 Corrupt images. CIs are defined as being caused by problems that invalidate the whole data set. They appear mostly in connection with a radar failure in either radar hardware or software, such as missing noise samples, buffer clearance or data transmission faults. During maintenance, for instance, radar sensitivity is increased in order to emphasize the low intensities. After the restart of operation, the first one or two scans sometimes keep these settings so that the intensity of the entire data set is too high. Hence, one type of CI is characterized by a high number of range bins with enhanced (noise) values evenly distributed over all ranges and azimuths (see figure 2(a)). In particular, a great many range bins at the most distant range kilometre contain non-zero values. Based on these characteristics, three test criteria have been specified: (i) mean intensity, (ii) percentage of non-zero range bins and (iii) fraction of relevant range bins at the most distant range kilometre. The data set is marked as CI if more than 67% of all range bins contain values, the mean intensity is lower than 10 dBZ and more than 75% of the most distant range kilometre contain values greater than 2.5 dBZ.

Another type of CI is particularly prone to cause problems in the assimilation process. It is characterized by highly variable reflectivities (also mentioned by Berenguer *et al.* (2006)) close to the radar site whereas range bins further away contain no reflectivity values (see figure 2(*b*)). The most probable explanation of this phenomenon seems to be a certain kind of strong anomalous propagation (anaprop). Reasons for anaprop have been well known for many years and are explained in detail in, for example, Alberoni *et al.* (2001) or Steiner and Smith (2002). The vertical profiles of atmospheric parameters such as pressure, water vapour partial pressure and air temperature influence the propagation characteristics of electromagnetic signals in the atmosphere at microwave frequencies. A measure for the propagation characteristics is the refractivity. If the refractivity is higher than expected from a standard atmosphere,

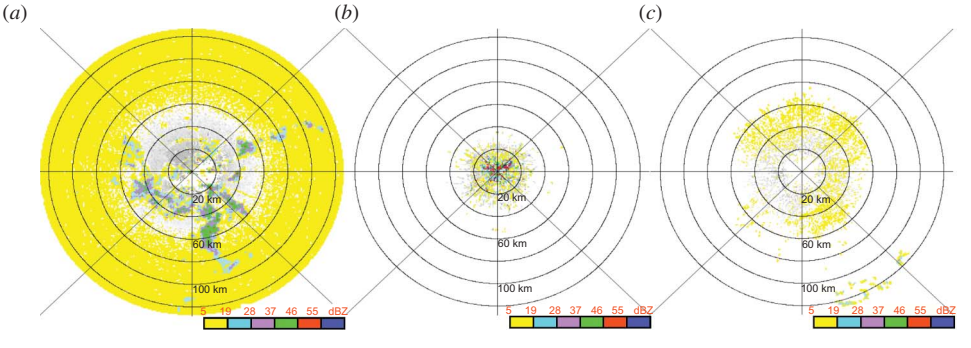

Figure 2. Two species of corrupt image (*a*) and (*b*) and 'German pancake' (*c*) cases within data sets of a precipitation scan of a German weather radar. In all three cases, the entire data set is flagged and not used for the assimilation process.

the radar rays may intersect the Earth's surface resulting in erroneous echoes. Under circumstances where anaprop is extremely strong, the detected signals are located close to the radar site and no range bins with reflectivity values can be detected further away (note that data from just one low elevation angle are analysed). Both Cho *et al.* (2006) and Berenguer *et al.* (2006) used a fuzzy logic approach to identify anaprop (and ground echoes) at Canadian and Spanish radar sites, respectively. Even though both methods show promising results, they are not applicable to the present 2D data as they are based on parameters available only from a volume scan. Harrison *et al.* (2000) described another approach for the detection of anaprop. Rather than analysing the reflectivities and radial velocities, the probability of precipitation is determined based on the analysis of synoptic information, although this was deemed impractical under near real-time conditions here.

This work uses the following textural detection algorithm instead. First, the number of non-zero range bins in the region beyond a range of 75 km (outer region) is checked. In case there exists almost no non-zero range bins in the outer region, basic statistics such as mean, standard deviation and variance are calculated for a test cluster ring of 4×4 range bins around the radar site. The range of every group remains the same (generally between 15 and 18 km (inner region)), whereas the azimuth angle is changed by 1°. This procedure allows the characterization of reflectivity variability between the small test clusters. However, in order to ensure the confidence level, at least 250 of the checked test clusters should contain more than eight non-zero range bins. Using this algorithm, the combination of the described inner and outer regions' characterization of a data set is very successful in detecting the phenomenon described above.

4.1.2 'German pancake'. The phenomenon we call the GP is characterized by its preferred appearance at radar sites in big cities (mostly during daytime and spring/summer/fall season); the confinement of the vertical extension of the reflectivity echoes to the lowest 2–3 km; the fact that many range bins close to the radar show low reflectivity (see figure 2(*c*)); and that the histograms of reflectivity show a narrow Gaussian distribution (mean width of about 5 dBZ). Although all GP cases identified at the German radar sites show the above-mentioned common characteristics, their definite meteorological conditions could not ultimately be established, as this phenomenon is relatively rare. During 4 years of precipitation scan data, only about 150 cases of GP appeared at the 16 radar sites in Germany.

In GP cases, reflectivities mostly show a circular shape around the radar site that changes size, but not position, during the day. This, together with the low vertical extension, provided the motive for the naming of this phenomenon. Although GPs do occur at all German radars, the frequencies at the Berlin and Hamburg sites are the highest, possibly because of special meteorological conditions, such as rising warm air packages (clear air echo; Atlas (1990)). Another explanation could be huge swarms of insects flying at the same time. These bio-meteors form a widespread pattern close to the radar with low intensity echoes (Peura 2002). Yet another explanation could be based on the scattering of the radar beam on dust particles or aerosol clusters in the vicinity of big cities.

Since vertical information (like volume scans) is available only every 15 min, and the time of day and the month of appearance are weak identification criteria, the GP identification algorithm in 2D data was based on the histogram of reflectivities. Tests have shown that GP reflectivity histograms differ significantly from cases of precipitation,

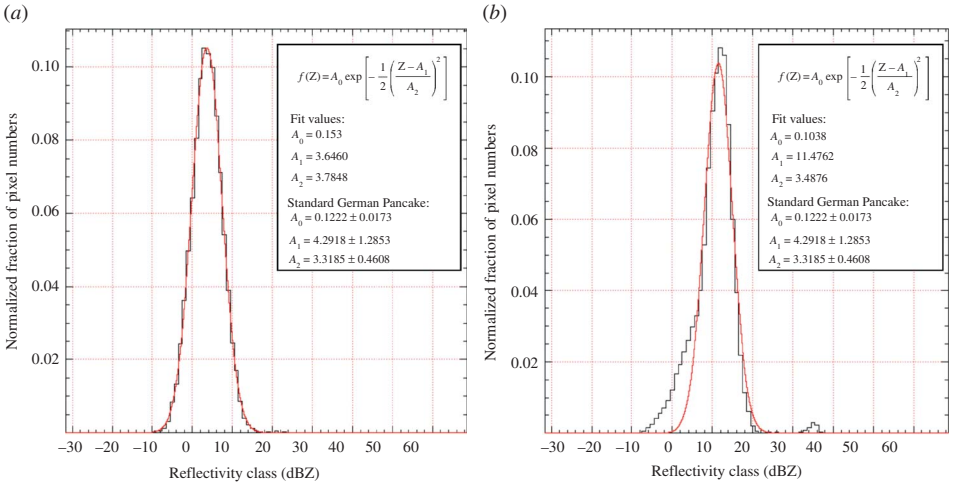

Figure 3. Histograms of reflectivity for an example of 'German pancake' (*a*) and stratiform precipitation (*b*) with their respective Gaussian curve fit in red. In case of real precipitation, the centre of the Gaussian fit is shifted to slightly higher reflectivity values.

both convective and stratiform. In both cases, the histogram is more widespread over several reflectivity classes than in cases of GP (see figure 3). In addition, the modal value of convective precipitation histograms occurs at higher reflectivities than in GP. More than 100 visually selected GP cases were analysed and a standard GP reflectivity distribution was established. It shows a very narrow peak at small dBZ values and almost no higher dBZ values (see figure 3). For each test case, the width and centre of a fitted Gaussian curve are determined. The mean of these two parameters are the main detection criteria for GP cases. Following this, histograms of every new radar data set are calculated and a Gaussian curve is fitted. In figure 3, histograms of a GP example and stratiform precipitation are shown with their respective Gaussian fit. Additionally, differences between the histogram and the Gaussian fit in the precipitation case are more pronounced, especially at the base of the Gaussian fit. After determination of the width and centre of the fitted Gaussian curve for every new radar image, the parameters of the fit are compared to the GP 'standard'. If the histogram parameters of the analysed data set lie beyond the double standard deviation area of the GP 'standard', the data field is considered to be problematic. The standard reflectivity distribution is chosen so that GP cases of all 16 radar stations can be detected, although, as noted above, most cases appear at the Berlin and Hamburg sites.

4.2 *Problems of individual range bins*

In the second part of the procedure, single range bin errors are detected. In this case, the appropriate bit is set for the corresponding range bins in the quality-index field.

4.2.1 Spokes and rings. Spokes are areas where the intensity changes suddenly over several range kilometres in the azimuthal direction. Shadowing effects due to obstacles like mountains, radio towers or other high buildings cause a decrease of intensity over a certain range as compared to the neighbouring azimuths (negative spokes). On the

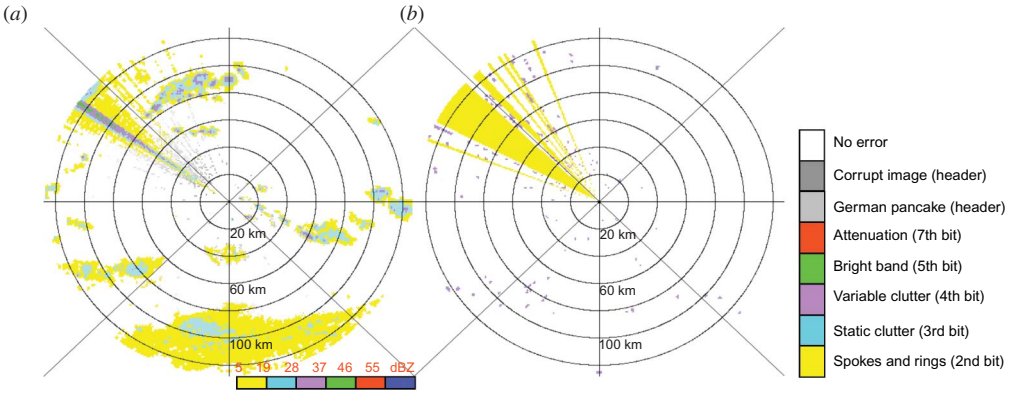

Figure 4. Radar image on 5 February 2007 13:40 UTC at the Hamburg radar (*a*) and the corresponding quality-index field (*b*). Bits 1, 6 and 8 are vacant. Detection algorithms for attenuation and bright band are under development.

other hand, external transmitters or sometimes internal technical problems (e.g. slow ray buffer) may cause an increase of intensity for a more or less wide range (positive spokes). The most frequent type of interfering radiation appears as a straight line segment as described in Peura (2002). In some cases, however, especially during military exercises, the segments are wider (see figure 4) or appear as a 'flower' around the radar site. For the detection of all these phenomena, their special geometric characteristics in polar coordinates are used. The search of (normally) sharp edges towards the adjacent azimuth is the most important feature for identifying spokes.

In order to assign the edges to the start and end rays of a spoke, respectively, the located edges are classified into positive (clockwise increasing intensity) and negative ones (clockwise decreasing intensity). For every pair of edges, some tests are necessary in order to state the identification of the spokes more precisely. Sometimes, there are several edges within a spoke, as in the example in figure 4. Then, the subsequently described check starts with the outermost located edges. If this pair of edges fails one of the tests, then the other pairs are checked in the same way. In the first step, the lengths of the edges are checked. This parameter determines the allowed width of the spokes (distance between edges), because in our algorithm wide spokes must have more distinct edges than narrow spokes. If the tested edge fails this check, the respective positive or negative edge is rejected. In the second step, the intensity difference between the spoke and the surrounding range bin values is tested. If the difference in the mean intensity between the spoke and the surrounding range bin values is lower than 30%, the respective positive or negative edge is rejected. The non-rejected edges after these tests are assumed to belong to real spokes and the whole area (all azimuths) between the non-rejected edges are marked in the quality-index field in a reserved bit (second bit, shown in yellow in figure 4(*b*)). If the area contaminated by spokes is very large, then the radar data set is additionally marked as CI in the header of the quality-index field.

In a B-scan (range over azimuth), this spoke-detecting algorithm is, in principle, also valid for rings. Ring segments are areas where the intensity at some ranges is considerably higher or lower compared to the adjacent ranges over a certain span of azimuths. This is often caused by ship radars or, rarely, by technical problems. For the detection

of rings the (normally) sharp edges towards the adjacent range are also searched. The located edges are checked with the tests mentioned above, but slightly different thresholds for the length–width criterion are used. The ranges between recognized edges are marked in the quality-index field in the same bit as the spokes (second bit).

4.2.2 Static and variable clutter. Even though there are Doppler clutter filters and thresholds applied to the data sets of the German weather radar network (Seltmann 2000), there may still be clutter remnants left in the data depending on the filter and threshold settings. In particular, wind parks and ships cannot be detected with a Doppler filter since the targets are moving. The clutter remnants appear as single range bins or small groups of range bins (clusters) with highly variable reflectivity values. Accordingly, two different types of clutter were defined. The first type is static clutter, for example, wind-parks, that show up in almost every data set at the same location. The second type is variable clutter, for example remnants of ground clutter, aircraft or ships, which are more or less randomly distributed over all ranges and azimuth angles.

The new detection algorithm deals with these clutter problems separately. Static clutter is marked by a blacklist. This idea is not new, but nevertheless it is one of the most effective (e.g. as suggested by Harrison *et al.* (2000)). Hannesen (2001) described a clutter map technique where radar reflectivities are sampled during fair weather and then subtracted from measured reflectivities during precipitation events. A combination of a static clutter map and a texture-based clutter elimination algorithm is described and tested by Gabella and Notarpietro (2002). In the approach described here, site-specific static clutter maps (blacklists) have been developed by analysing over 200 precipitation-free data sets. Range bins with a precipitation probability exceeding a certain threshold (currently 40%) have been entered into the blacklist. Then, for each current data set, the coordinates of non-zero range bins are compared to the blacklist. Whenever positions coincide, the range bin is marked as 'clutter' in the quality-index field in a separate bit (third bit, shown in cyan in figure 4(*b*)). Since new wind parks or other echo sources may appear, it is necessary to update the blacklist automatically on a regular basis, as was pointed out by Rico-Ramirez *et al.* (2007). To this end, precipitation-free data sets are continuously searched during quality check and, if appropriate, stored in a ring archive. Blacklists are then renewed on a monthly basis for each radar site.

The second part of the clutter detection algorithm deals with variable clutter. The removal of ground-, sea- and anaprop clutter has been reviewed by many researchers (Berenguer *et al*. 2006, Cho *et al*. 2006, Rico-Ramirez and Cluckie 2008). Again, a simple and robust test is applied here. It is based on the fact that real precipitation spreads over some minimum area and is, in the overwhelming majority of cases, not an event with only few range bins involved. Thus, all range bins with a non-zero reflectivity are checked for a connection to other non-zero range bins in the first step. In the second step, connected range bins (or clusters) are separately tested against two maximum thresholds: the number of range bins connected within the cluster, and the number of range bins with non-zero reflectivities surrounding the cluster (similar to the approach described in Fulton *et al.* (1998)). With the first threshold, the minimum spatial extension of real precipitation is taken into account. Clusters exceeding the threshold are not considered to be clutter. With the second threshold, the proximity of the cluster to other clusters is tested. If the maximum allowed number of non-zero range bins in the surrounding of the cluster is exceeded, the cluster is not considered to be clutter.

The surrounding is defined in three range-dependent classes to ensure an even spatial extension with increasing range. The cluster is flagged to be a variable clutter in the quality-index field only if both criteria are met (fourth bit, shown in purple in figure 4(*b*)).

4.3 *Tests and evaluation of the algorithms*

The methods described above were tested over nearly 2 years in a quasi-operational environment by inspection and subjective judgement by radar experts. The results of these tests are summarized in the following subsections and table 1.

At the start of our work, several years of routinely measured precipitation scans were available from all 16 German radar sites. Those scans were the basis for the development of all described algorithms. First, many precipitation scans were visually categorized to distinguish between scans with realistic precipitation and erroneous signals. Only about two-thirds of the selected scans for each category of erroneous signals (CI, GP, rings and spokes, static and variable ground clutter) were then used as the basis for the development of the particular detection algorithm (training data set). The remaining third of the selected scans was used as a validation data set. Both the training and the validation data set were constantly growing since the visual categorization of the daily, newly performed precipitation scans was continued during the development of the detection algorithms. The newly developed algorithms were tested against selected cases that had not been used for their development.

The number of test cases for each method were divided into H 'hits' (all correct error detections), F 'false alarms' (all incorrect detections), M 'missed detections' (all incorrectly passed errors) and Z (all correct passes). Two verification scores were computed from these four values. The probability of detection (POD) or hit rate, defined as $HR = H/(H+M)$, measures the fraction of observed radar errors that were correctly detected. The probability of false detection, $POFD = F/(Z+F)$, also known as the false alarm rate, gives the fraction of incorrect detections. A comparison with independent data, for example, rain gauge data, has not been attempted due to the different spatio-temporal structures of the measurement techniques and because a correction of the detected radar errors would be necessary, which is beyond the scope of this article.

4.3.1 Corrupt images. The relation between hits and false alarms may be adjusted by the choice of test thresholds. The difficulty in the definition of the CI thresholds described in §4.1.1 is to distinguish light rainfall from CIs. The hit rate is 0.95 with the thresholds given above. These statistics are not significant yet, because CI do not occur very often, so that our validation data set comprises only 20 cases. However, the false alarm rate of 0.02 is negligible. In such cases, a comparison with data previously

Table 1. Summary of verification.

Radar error	Hit rate	False alarm rate
Corrupt image	>0.95	<0.02
German pancake	0.90	0.01
Spokes and rings	0.80	<0.20
Static and variable clutter	>0.70	<0.15

received from the particular radar and also with data from the adjacent radars in the overlapping areas could help to avoid misinterpretations, as suggested by Harrison *et al.* (2000).

The second part of the algorithm was tested on about 20 cases and was able to detect all of them, resulting in a hit rate of one. Furthermore, not a single precipitation case has been erroneously assigned to this type of CI, so that the false alarm rate is 0. However, this phenomenon occurs even less often than GP (see §4.1.2). Therefore, the detection capability of the algorithm cannot be considered to be sufficiently tested nor ultimately proved, yet.

4.3.2 German pancake. With a false alarm rate of 0.01 only very few precipitation data fields are erroneously marked as GP. Since GP conditions do not occur very often either, statistics about correctly marked data fields are still rare. About 50 known GP cases, which were not included in the standard calculation, were tested with a hit rate of 0.9.

4.3.3 Spokes and rings. The test over 2 years and 900 cases showed that most of the spokes and rings are detected correctly, which is reflected in a hit rate of 0.8. In contrast, the false alarm rate differs for spokes and rings. It is as low as 0.1 for spokes, because only in rare cases was radially shaped precipitation misdetected. The false alarm rate for rings is 0.2, because rings are more often detected in the noise near the radar. Depending on the use, this may be acceptable and is mainly a question of the thresholds applied. For NWP data assimilation, small and less intensive spokes and rings should be reliably identified, so the present thresholds were set rather aggressively.

4.3.4 Static and variable clutter. The two parts of clutter identification in the algorithm were tested in about 50 cases for each radar. The hit rate of marked range bins in precipitation-free data fields was determined to range between 0.9 and 1, depending on the typical radar site clutter characteristics. In cases where data fields contain precipitation, not all clutter range bins could be detected, while, on the other hand, some small rain cells were flagged. In this case, the hit rate is 0.7 and the false alarm rate is 0.15. The misinterpretation is mainly because small rain showers can have a maximum spatial extension that is very similar to the extension chosen here for the identification of ground clutter remnants. At this first stage of operational quality control, the trade-off between errors of the first (misses) and second kinds (false alarms) is accepted to enable a successful data assimilation. At the 3D stage to follow (see outlook in §6), further development and improvement will be carried out. With respect to clutter and speckles, the vertical extent will be accounted for.

5. Creation of a quality-controlled composite

One of the main questions in the creation of a radar composite is to decide which range bins should be used in the overlapping areas of several radars. Historically, the range bin with the highest reflectivity value had been accepted for the German radar composites. Thus, gross radar errors have been emphasized, as shown in figure 5(*c*).

Table 2. Priority list for the different types of radar errors.

Type of error	Radar error	Priority
Gross errors affecting the total data field	Corrupt image	2
	'German pancake'	3
Errors affecting the occurrence of precipitation	Spokes and rings	4
	static clutter	5
	variable clutter	5
Errors affecting the amount of precipitation	Attenuation	6
	bright band	6
Error free	No error	20

In order to utilize the quality information from each radar station in the construction of a composite, new rules of composite generation had to be developed (Helmert and Hassler 2006). First, a priority list has to be defined for the different types of radar errors. Table 2 shows the actual version of the priority list for use in COSMO-DE. Sensitivity studies have shown that large-area faulty information and static clutter have the highest negative impact on data assimilation. Furthermore, it has been shown that it is more important to know if precipitation occurred at all than its exact amount (Leuenberger 2005; Stephan *et al*. 2008). Therefore, gross errors (e.g. CI or GP) are assigned lowest priorities, whereas errors affecting the occurrence of precipitation (e.g. radial spokes, clutter) are given a medium value. Errors affecting the amount of precipitation (e.g. attenuation, bright band) receive a higher priority than the errors mentioned before. Error-free range bins get the highest priority compared to flagged range bins. The priority list sets up one possible quality measure by a weighted combination of error flags. It can easily be modified for different applications.

The ranking information is used in the creation of the radar composite for data assimilation. Again, both a precipitation composite and an additional quality-index composite are generated. In the overlapping areas of several radars, the range bin with the highest priority in conformity with the defined priority list is accepted for both composites. If the priorities of the range bins are equal, or if the quality-index field is unavailable from one of the radar stations, then the range bin with the highest reflectivity value is used. In the case that no quality-index field is available, a missing value is written in the quality-index composite. The same approach is used if multiple range bins are inside one composite grid point. Figure 5(*a*) and (*b*) shows an example of the precipitation composite and the quality-index composite. In comparison to the former maximum method (figure 5(*c*)), the occurrence of spokes and clutter in the overlapping areas could clearly be reduced. The remaining artefacts are flagged in the quality-index composite (figure 5(*b*)).

At this stage, for assimilation in COSMO-DE, all flagged range bins in the quality-index composite are set to 'no data' in the precipitation composite as shown in figure 5(*d*). Thus, the detected spokes at the radar stations of Hamburg, Hannover and Emden (all in northwest Germany) are removed along with most of the clutter remnants. The distinction between range bins with no radar information, that is no measurement or range bins removed by quality control (white colour in figure 5), and range bins with no measureable precipitation (values lower than 0.01 mm hour^{-1} – grey colour in figure 5) is an important indicator for the data assimilation.

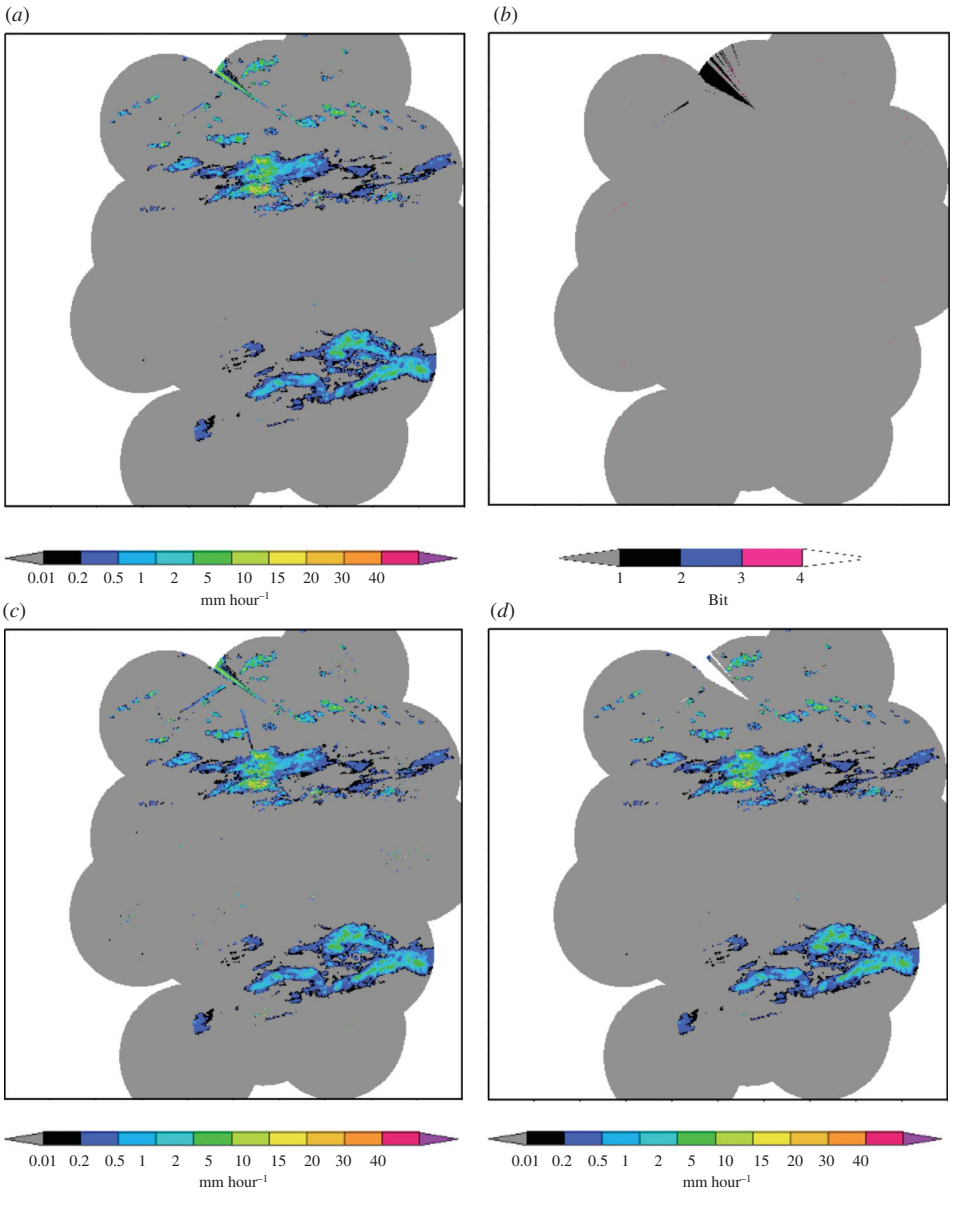

Figure 5. Radar composite on 5 February 2007 at 13:40 UTC (*a*) using quality information for assigning range bin values in the overlapping areas (mm hour^{-1}); (*b*) quality-index composite (Bit); (*c*) using the maximum method for assigning range bin values in the overlapping areas (mm hour^{-1}); (*d*) all flagged range bins in the quality-index composite are set to 'no data' (white colour) (mm hour^{-1}).

6. Summary and outlook

The quality control procedure described in this article consists of several modules that are combined to produce a chain of error identification steps. First, the data fields are searched for gross general problems such as GP or CIs. Second, single range bins or

groups of range bins are checked to determine whether they belong to spokes or rings or to stand-alone clusters of speckles. Each identified error is then flagged in a bit of the quality-index field. One advantage of detecting all errors separately prior to correcting them is that problematic range bins may be excluded from the determination of the correction value (e.g. clutter range bins in the vicinity of spokes for a spoke correction). It further enables a differentiated decision whether the range bins of the radar data set are to be used for a chosen application. Users preferring a precipitation confidence value between 0 and 1 may compute their application-specific probability for each range bin from the quality-index field. This concept can easily be used to create application-specific best radar data sets.

The primary result of the developments described in this article is the creation of a quality-index field along with the original data that allows the user to decide whether this particular interference should be utilized, excluded or corrected in follow-up algorithms. It is believed that this approach will be applicable to operational radars around the world, wherever robust real-time detection of interference is more essential than sophisticated individual correction schemes. Particularly, operational networks that are designed to satisfy many different kinds of users with diverging data quality aspects, such as the clients of national meteorological and hydrological services, may profit from the approach described here.

In order to take the quality information from each radar station into the composite, a quality-index composite is created along with the precipitation composite. In the overlapping areas of more than one radar, the range bin with the higher confidence value according to a user-defined priority list (or, if equal, with the higher precipitation value) is used. So far, for the assimilation in COSMO-DE, all range bins flagged in the quality-index composite are set to 'no data' in the precipitation composite. This way, the radar data quality for the latent heat nudging could be increased. Because the German radar network does not cover the entire model area of COSMO-DE, quality-assessed radar data from neighbouring countries are included as well.

Further applications will profit from the versatile concept of individual error flags. For instance, DWD's very successful cell tracking and warning system CONRAD (Lang 2001) has lately been adapted to include the information of the quality-index field. Thus, permanently detected wind park echoes and positive spokes are excluded from cell detection. This feature is still in the testing phase and not yet operational.

With the application of the procedure presented here in operational mode throughout the German radar network, the amount of spurious precipitation signals could be reduced significantly. This is reflected in hit rates between 0.95 and 0.70 for the individual detection algorithms. In some cases not all problematic range bins have been identified, whereas in other cases false alarms were issued. The algorithms are still being improved and will be extended by additional ones, such as bright band and attenuation detections.

This system is a very basic tool based on simple and robust algorithms. It is limited to flagging of a choice of errors in 2D reflectivity data only. No attempt has been operationally implemented so far to correct for errors detected, such as azimuthally interpolating negative spokes, or supplying a quantitative correction of the vertical reflectivity profile.

An extension of the concept towards volume radar data is currently underway. The system is being centralized to a new platform and extended to include radial wind data in addition to reflectivity. New flags, such as attenuation and folding are being implemented and some of the current algorithms improved. The new 3D tool overcomes some of the above shortcomings of the 2D tool presented here and is still more flexible.

It is currently being tested and will be presented in a follow-up paper. Furthermore, DWD has started to replace all the German network radars by dual polarization Doppler radars (to be finished by 2014). In this context, a new scan strategy will be introduced and the concept of a flexible combination of quality flags as presented here will be extended to dual polarization data.

Acknowledgements
The authors thank the two reviewers for their very helpful and constructive comments. Furthermore, they thank Paul J. Young and Todd J. Sanford for their corrections concerning the English style and grammar.

References

Alberoni, P.P., Anderson, T., Mezzasalma, P., Michelson, D.B. and Nanni, S., 2001, Use of the vertical reflectivity profile for identification of anomalous propagation. *Meteorological Applications*, **8**, pp. 257–266.

Atlas, D., 1990, *Radar in Meteorology*, 806 pp. (Boston, MA: American Meteorological Society).

Baldauf, M., Stephan, K., Klink, S., Schraff, C., Seifert, A., Förstner, J., Reinhardt, T. and Lenz, C.-J., 2007, The new very short range forecast model COSMO-LMK for the convection-resolving scale. *WGNE Blue Book*. Available online at: http://www.cmc.ec.gc.ca/rpn/wgne/

Berenguer, M., Sempere-Torres, D., Corral, C. and Sanchez-Diezma, R., 2006, A fuzzy logic technique for identifying nonprecipitation echoes in radar scans. *Journal of Atmospheric and Oceanic Technology*, **23**, pp. 1157–1180.

Cho, Y., Lee, G., Kim, K. and Zawadzki, I., 2006, Identification and removal of ground echoes and anomalous propagation using characteristics of radar echoes. *Journal of Atmospheric and Oceanic Technology*, **23**, pp. 1206–1222.

Doviak, R.J. and Zrnic, D.S., 1993, *Doppler Radar and Weather Observations* (San Diego, CA: Academic Press).

Fabry, F., Austin, G. and Tees, D., 1992, The accuracy of rainfall estimates by radar as a function of range. *Quaterly Journal of the Royal Meteorological Society*, **118**, pp. 435–453.

Friedrich, K., Hagen, M. and Einfalt, T., 2006, A quality control concept for radar reflectivity, polarimetric parameters and Doppler velocity. *Journal of Atmospheric and Oceanic Technology*, **23**, pp. 865–887.

Fulton, R.A., Breidenbach, J.P., Seo, D., Miller, D.A. and O'Bannon, T., 1998, The WSR-88D rainfall algorithm. *Weather and Forecasting*, **13**, pp. 377–395.

Gabella, M. and Notarpietro, R., 2002, Ground clutter characterization and elimination in mountainous terrain. In *Proceedings of the Second European Conference on Radar in Meteorology and Hydrology*, 18–22 November 2002, Delft, The Netherlands (Göttingen: Copernicus GmbH), ERAD Publication Series 1, pp. 305–311.

Golz, C., Einfalt, T. and Galli, G., 2006, Radar data quality control methods in VOLTAIRE. *Meteorologische Zeitschrift*, **15**, pp. 497–504.

Hannesen, R., 2001, *Quantitative Precipitation Estimation from Radar Data – a Review of Current Methodologies*. Deliverable MUSIC Project Report No. 4.1 (Neuss: Gematronik GmbH). Available online at: http://www.einfalt.de/literature/Hanne_Quant.pdf

Harrison, D.L., Driscoll, J. and Kitchen, M., 2000, Improving precipitation estimates from weather radar using quality control and correction techniques. *Meteorological Applications*, **7**, pp. 135–144.

Hassler, B., Helmert, K. and Seltmann, J., 2006, Identification of spurious precipitation signals in radar data. In *Proceedings of the Fourth European Conference on Radar in Meteorology and Hydrology*, 18–22 September 2006, Barcelona, Spain (Göttingen: Copernicus GmbH), ERAD Publication Series 3, pp. 590–592.

Hassler, B., Wagner A., Seltmann, J. and Lang, P., 2005, *The German Pancake – a Radar Mystery*. EGU General Assembly Abstracts, No. EGU05-A-08117 (Göttingen: Copernicus GmbH).

Helmert, K. and Hassler, B., 2006, Development and application of a quality-index composite. In *Proceedings of the Fourth European Conference on Radar in Meteorology and Hydrology*, 18–22 September 2006, Barcelona, Spain (Göttingen: Copernicus GmbH), ERAD Publication Series 3, pp. 587–589.

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**, pp. 269–277.

Joss, J. and Lee, R., 1995, The application of radar-gauge comparisons to operational precipitation profile corrections. *Journal of Applied Meteorology*, **34**, pp. 2612–2630.

Joss, J. and Waldvogel, A., 1990, Precipitation measurement and hydrology. In *Radar in Meteorology*, D. Atlas (Ed.), pp. 577–606 (Boston, MA: American Meteorological Society).

Klink, S. and Stephan, K., 2004, Assimilation of radar data in the LM at DWD. *COSMO Newsletter*, **4**, pp. 143–150. Available online at: www.cosmo-model.org

Laborda, Y., Salles, C. and Neppel, L., 2006, Influences of radar rainfall uncertainties on a distributed rainfall runoff model. In *Proceedings of the Fourth European Conference on Radar in Meteorology and Hydrology*, 18–22 September 2006, Barcelona, Spain (Göttingen: Copernicus GmbH), ERAD Publication Series 3, pp. 426–428.

Lang, P., 2001, Cell tracking and warning indicators derived from operational radar products. In *Preprints, 30th International Conference on Radar Meteorology*, Munich, Germany (Boston, MA: American Meteorological Society), pp. 245–247.

Leuenberger, D., 2005, High-resolution radar rainfall assimilation: exploratory studies with latent heat nudging. PhD thesis, 15884, ETH Zurich, Zurich, 103 pp.

Marx, A., Kunstmann, H., Bardossy, A. and Seltmann, J.E.E., 2006, Radar rainfall estimates in an alpine environment using inverse hydrological modelling. *Advances in Geosciences*, **9**, pp. 25–29.

Michelson, D.B. and Sunhede, D., 2004, Spurious weather radar echo identification and removal using multisource temperature information. *Meteorological Applications*, **11**, pp. 1–14.

Peura, M., 2002, Computer vision methods for anomaly removal. In *Proceedings of the Second European Conference on Radar in Meteorology and Hydrology*, 18–22 November 2002, Delft, The Netherlands (Göttingen: Copernicus GmbH), pp. 312–317.

Rico-Ramirez, M.A. and Cluckie, I.D., 2008, Classification of ground clutter and anomalous propagation using dual-polarization weather radar. *IEEE Transactions on Geoscience and Remote Sensing*, **46**, pp. 892–1904.

Rico-Ramirez, M.A., Cluckie, I.D., Shepherd, A. and Pallot, A., 2007, A high resolution radar experiment on the Island of Jersey. *Meteorological Applications*, **14**, pp. 117–129.

Seltmann, J.E.E., 2000, Clutter versus radar winds. *Physics and Chemistry of the Earth*, **B25**, pp. 1173–1178.

Seltmann, J.E.E. and Riedl, J., 1998, Improved clutter treatment within the German radar network: first results. In *COST-75 Final Seminar*, Locarno, Switzerland (Brussels: European Commission), pp. 267–279.

Steiner, M. and Smith, J., 2002, Use of three-dimensional reflectivity structure for automated detection and removal of nonprecipitating echoes in radar data. *Journal of Atmospheric and Oceanic Technology*, **19**, pp. 673–686.

Stephan, K., Klink, S. and Schraff, C., 2008, Assimilation of radar derived rain rates into the convective-scale model COSMO-DE at DWD. *Quarterly Journal of the Royal Meteorological Society*, **134**, pp. 1315–1326.

Sun, J. and Wilson, J.W., 2003, The assimilation of radar data for weather prediction. *Meteorological Monographs*, **30**, pp. 175–198.