University of Reading

The potential use of operational radar network data to evaluate the representation of convective storms in NWP models

Article

Published Version

Creative Commons: Attribution-Noncommercial 4.0

Open Access

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/qj.3793

Publisher: Royal Meteorological Society

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

RESEARCH ARTICLE

The potential use of operational radar network data to evaluate the representation of convective storms in NWP models

Thorwald H. M. Stein[1] | Robert W. Scovell[2] | Kirsty E. Hanley[3] | Humphrey W. Lean[3] | Nicola H. Marsden[1,2]

[1]Department of Meteorology, University of Reading, UK

[2]Met Office, Exeter, UK

[3]MetOffice@Reading, Reading, UK

Correspondence
T.H.M. Stein, Room 507A, Lyle Building, Whiteknights, PO Box 243, Reading RG6 6BX, UK
Email: t.h.m.stein@reading.ac.uk

Abstract

Operational forecasting centres increasingly rely on convection-permitting NWP simulations to assist in their forecasting of convective events. The evaluation of upgrades in the underlying NWP modelling system normally happens through routine verification using traditional metrics on two-dimensional fields, such as gridded rainfall data. Object- and process-based evaluation can identify specific physical mechanisms for model improvement, but such evaluation procedures normally require targeted and expensive field campaigns. Here, we explore the potential use of the UK operational radar network observations and its derived 3D composite product for evaluating the representation of convective storms in the Met Office Unified Model. A comparison of the $1 \times 1 \times 0.5$ km 3D radar composites against observations made with the research-grade radar at Chilbolton in the southern UK indicates that the 3D radar composite data can reliably be used to evaluate the morphology of convective storms. The 3D radar composite data are subsequently used to evaluate the development of convective storms in the Met Office Unified Model. Such analysis was previously unavailable due to a lack of 3D radar data of high temporal frequency. The operational nature of the UK radar data makes these 3D composites a valuable resource for future studies of the initiation, growth, development, and organisation of convective storms over the UK.

KEYWORDS

convective storms, echo-top height, model evaluation, radar meteorology, storm morphology

1 | INTRODUCTION

The gradual reduction in numerical weather prediction (NWP) model grid size as computational resources expand permits an increasing complexity in the physical processes that are simulated. For an NWP model with grid size of the order of 1 km, individual convective storms can be represented explicitly, that is, without requiring a parametrization scheme to represent the effects of deep convection on sub-grid scales. Many improvements in forecasting of rainfall from convection have been attributed to the explicit representation of convection, including a better simulation of the diurnal cycle of convection in the Tropics (e.g., Pearson *et al.*, 2010) and a higher accuracy in predicting the location of intense rainfall (Roberts and Lean, 2008). Invariably, due to the partial representation of smaller-scale processes such as turbulent mixing, errors and biases in rainfall persist or new ones are introduced, even at smaller grid lengths (Hanley *et al.*, 2015; Clark *et al.*, 2016). To further improve and verify these models, a process-based evaluation of convection in NWP models typically requires focused and extensive field campaigns, using multiple observation platforms. However, Stein *et al.* (2015) performed object-oriented model evaluation in combination with a composite statistical approach, providing a useful overview of the representation of convective storms in NWP models using only observations from a single steerable research-grade radar. For ongoing model development, however, an evaluation procedure making best use of readily available observations will allow on-demand evaluation of cases of interest as these arise. It is therefore worth exploring whether operational radar network data can provide equally useful results to research-grade radar observations for model evaluation of the characteristics of convective storms.

In 2011–2012, more than 1,000 convective storms over the southern UK were scanned with the Chilbolton Advanced Meteorological Radar (CAMRa) during the DYMECS project (Dynamical and Microphysical Evolution of Convective Storms; Stein *et al.*, 2015). CAMRa's 0.28 ° beam width enables observations of cloud microphysical and dynamical processes at scales of less than 300 m out to 100 km range, sufficient for the evaluation of high-resolution NWP models. By analysing storm characteristics based on statistics over many observed storms, the DYMECS project demonstrated that the Met Office Unified Model (MetUM) versions in use at the time generated convective storms that are a factor 1.5–2 wider than observed when run at a 1,500 m grid length, while storms of comparable size to observations are generated at grid lengths of 200 m or less (Stein *et al.*, 2014). Using the Doppler measurements of CAMRa, a statistical

evaluation of updraughts indicated that the MetUM at grid lenghts of 200 and 100 m generates updraughts that are of comparable strength to the retrievals, but slightly narrower than observed (Nicol *et al.*, 2015). These object-oriented evaluation studies are instrumental in steering model development and improvement. However, CAMRa's primary purpose as a research radar (i.e., it is not run operationally nor with a set scanning procedure) and its low scanning velocity mean that it cannot be used for the on-demand model evaluation envisaged here.

Typically, operational meteorological radars are operated using a scanning strategy that targets precipitation, so that rainfall and snowfall rate retrievals can be obtained operationally for advance warning of flooding and related severe weather (Fabry, 2015). Thus, model evaluation using radar network data has focused on spatial variability of rainfall (Roberts and Lean, 2008), or rainfall feature characteristics (Davis *et al.*, 2006). Recently, however, Scovell and Al-Sakka (2016) presented 3D radar composite data for small domains over France and the UK as a means to assist air traffic control with locating severe convection. Such a data product could in principle be available for on-demand model evaluation. Given that the typical weather radar beam width is 1°, the smaller scales observed by CAMRa cannot be analysed with such a 3D composite. However, the operational availability and the extended coverage of the UK radar network compared to CAMRa means that storms can be analysed and tracked for model evaluation at all times.

The purpose of this paper is therefore to explore the use of the 3D composite data for evaluating convective storm characteristics in convection-permitting NWP models. This paper will primarily focus on convective storm morphology, that is, a macrophysical storm characteristic, following the methodology and results of Stein *et al.* (2014). In Section 2, the UK radar network is presented, including a description of the 3D compositing method following Scovell and Al-Sakka (2016), as well as the CAMRa observations and the MetUM simulations generated in the DYMECS project. Storm characteristics derived from the 3D composite data are compared against CAMRa in Section 3 to assess whether the results from Stein *et al.* (2014) can be reproduced with the 3D composite data. In Section 4, storm morphologies derived from the 3D composite data are used to evaluate MetUM simulations, to demonstrate the use of the 3D composite data to analyse the statistics of many storms simultaneously, which was not possible with the CAMRa observations in DYMECS. Recommendations for future use of operational radar network data for model evaluation are provided in the conclusions in Section 5.
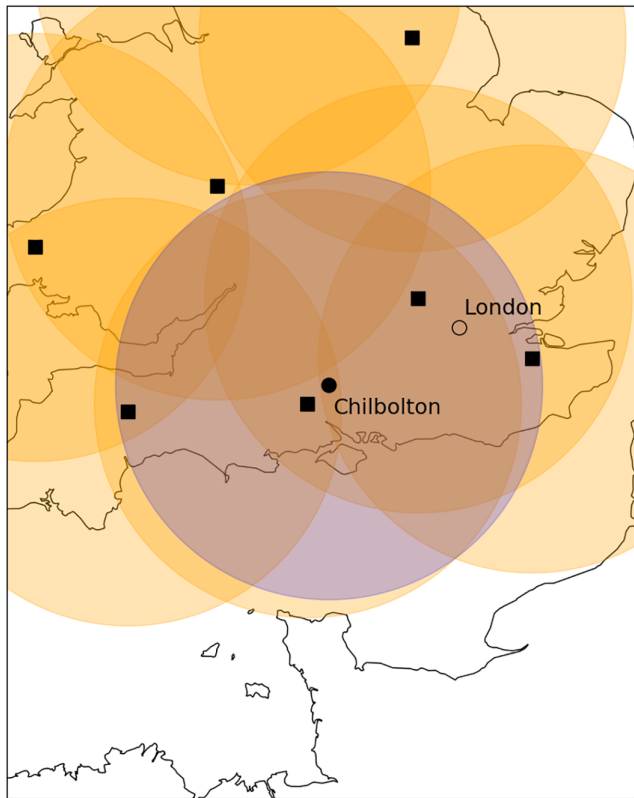
FIGURE 1 Map of the southern UK centred on Chilbolton (black dot). Individual radar locations are indicated by black squares, surrounded by 150 km range rings for the network radars (orange) and CAMRa (blue)

2 | DATA AND METHODOLOGY

For the evaluation of the Met Office 3D composites, four days were selected from the DYMECS cases for which many RHI scans were performed through storms, covering a range of cloud-top heights as identified using a 0 dBZ threshold. In this paper, we will refer to these heights as echo-top heights (ETH), to more precisely refer to the maximum height at which a particular radar reflectivity value or "echo" is observed. The four days include two summer days with deep convection, that is, 0 dBZ ETH above 8 km, namely 26 August 2011 and 25 August 2012; and two spring days with 0 dBZ ETH typically between 4 and 6 km, namely 12 April 2012 and 6 June 2012. All days were dominated by scattered convective storms in the early afternoon, organising into rainbands towards the end of the day.

2.1 | UK radar network

The UK operational radar network consists of 18 C-band (5 cm wavelength) radars located in the United Kingdom and Ireland and on Jersey. For the present study, only the radars at Chenies, Clee Hill, Cobbacombe, Dean Hill,

and Thurnham are relevant, as measurements from these radars contribute to the 3D composite centred around Chilbolton (Figure 1). The radar scan strategy involves a number of plan-position indicator (PPI) scans at low elevations, up to 4° in elevation, targeting precipitation; the PPIs are averaged to 600 m range gates and 1° in azimuth, close to the radar beam width of 1.1°. The entire scan cycle takes up to 5 min and the primary operational product of these observations is a 5 min rainfall rate product covering the entire UK and Ireland (Harrison *et al.*, 2000).

For the purpose of this study, we are interested in radar observations within 100 km range of Chilbolton. For all land locations within this range, at least one UK radar is within 100 km distance, so that every radar sample volume in this domain will have a maximum diameter of 1.75 km. Combining observations from all available UK radars within 100 km range of Chilbolton, at least six scans are available per 1 km^3 at 250 m above mean sea level, while at 11,750 m, at least three scans are available (Figure 4 in Scovell and Al-Sakka (2016)). Between these two extreme heights, the number of available observations rises to a maximum of 19 for some areas of the domain.

2.1.1 | 3D composite retrieval

Scovell and Al-Sakka (2016) describe several procedures to merge the radar measurements on to a Cartesian grid, which are, in summary,

(1) a search algorithm that finds all observations points x_b within a given radius $R_B = 3.0$ km (horizontal distance) of the Cartesian grid point x_i; and

(2) a distance-weighted average of Z (in dBZ units, rather than mm$^6 \cdot$ m^{-3}).

Two versions were considered in that study, namely one based on the Barnes (1964) weighting (the Barnes composite) and one based on the Zhang *et al.* (2005) weighting (the Zhang composite). For each grid point, the Zhang composite method finds, for each elevation in a volume scan, the nearest neighbour radar point within the search radius. These nearest-neighbour points are then interpolated vertically, for each radar, to get single-radar estimates of the dBZ at each grid point. The single-radar estimates are then weighted together, in dBZ, using a Gaussian distance-weighted mean. Conversely, the Barnes composite treats all points within the search radius equally and computes a weighted sum of dBZ, with weights based solely on a Gaussian function of each 3D radar-point-to-grid-point distance. Further, additional correction passes are made to the Barnes analysis, with progressively narrower Gaussian weight functions, to retrieve the fine-scale detail. In practice, it appears that the Zhang method retrieves better the peak dBZ values that are seen

in the raw data because the resulting estimates tend to be less smoothed, even taking into account the Barnes correction passes. Prior to calculating the composites, temporal synchronization is applied to the observation points using horizontal motion vectors derived from the previous two 3D radar composites.

The algorithms have since been further developed to mask and correct for terrain and beam blocking. The Cartesian pixels in the 3D radar mosaic are masked out when their centroids are below the ground height, with ground height estimated using satellite-derived digital terrain model data. The radar beam is considered blocked if the height of the centre of the beam minus half the beam width intersects the ground, with height computed using the approximation of the effective radius of the Earth equal to 4/3 the true radius of the Earth. A binary map is thus generated in azimuth and range for each elevation. Along each beam, at each range gate, the beam blocking fraction, f, is calculated as the fraction of blocked gates between the radar and the current gate. If a range gate has $f \geq 0.5$, it is marked as no-data and therefore not used in the 3D composite at all. If $f < 0.5$, the reflectivity value Z is divided by a factor $1 - f$. These corrections have very little impact at elevations greater than 1°.

At C-band, attenuation from heavy rainfall can lead to severe underestimates of the reflectivity values. At the time of production of these composites, an attenuation correction was applied based on Hitschfeld and Bordan (1954):

$$R = (Z/200)^{1/1.6}, \tag{1}$$

$$A = 0.0044R^{1.17}, \tag{2}$$

with reflectivity Z in mm$^6 \cdot$ hr^{-1}, rainfall R in mm·hr^{-1} and attenuation A in dB·km^{-1}. The attenuation correction was capped at 4.77 dB (a factor 3 in linear Z units). Since November 2016, a new attenuation correction scheme has become operational, which makes use of several independent estimates, described in section 3.2 of Dance *et al.* (2019).

2.2 | CAMRa observations during DYMECS

The Chilbolton Advanced Meteorological Radar (CAMRa) is an S-band (3 GHz) radar with a 25 m diameter antenna, leading to a narrow beam width of 0.28°, located in the southern UK, near Winchester (51°09'N, 1°26'W). The radar is calibrated with an uncertainty of less than 0.5 dB (Goddard *et al.*, 1994). The large antenna means that the

scan velocity is restricted to 2°s^{-1} and the radar is therefore not a suitable instrument to perform full-volume scans at the frequency required for studying convective storms. An automated track-and-scan algorithm was therefore designed in the DYMECS project, allowing volume scans for limited azimuth range to target a small number of convective storms. Using this algorithm, observations with CAMRa were collected on 40 days during 2011–2012. Individual convective storms were identified in the UK radar rainfall composite (Harrison *et al.*, 2000) using a threshold of 4 mm·hr^{-1} and were tracked over the course of their lifetime. The DYMECS radar scan strategy targeted cores of high radar reflectivity within these rainfall features, scanning vertically through these cores with multiple range–height indicator (RHI) scans, followed by sector plan-position indicator (PPI) scans, scanning horizontally across a group of convective storms at multiple elevations for storm volume reconstructions (Stein *et al.*, 2015). Such a scan cycle would typically take up to 15 min.

2.2.1 | Interpolation of CAMRa to 3D composite grid

We firstly wish to establish the accuracy of ETH identification in the radar composite. We consider three radar reflectivity thresholds to determine ETH, namely 0 dBZ (ETH$_0$), 10 dBZ (ETH$_{10}$), and 18 dBZ (ETH$_{18}$). The 0 dBZ threshold compares to a minimum detectable signal at 100 km of –3.5 dBZ for CAMRa and +0.5 dBZ for the UK radars. The 10 dBZ threshold is currently considered for use in model evaluation by the Met Office. The 18 dBZ threshold is chosen as it was used by Miltenberger *et al.* (2018) to study deepening of convective cells in the radar composite and model simulations.

To establish the accuracy of ETH identification in the radar composite data, RHI scans from CAMRa are used. For each radar composite, all RHI scans from the previous 5 min were considered. Since the DYMECS scanning strategy included multiple RHI scans along the same azimuth, if more than one scan occurred at the same azimuth during this 5 min period, only one scan was randomly selected for this azimuth to be included in the sample. The total number of scans in the sample for each day of interest is listed in Table 1.

Each RHI scan was linearly interpolated (in dBZ units) to the 1 km horizontal and 500 m Cartesian grid, x_i, used in the composite. An example of an interpolated RHI scan is shown in Figure 2. The interpolated CAMRa data retain features at the grid scale (1 km by 500 m) whereas the Barnes and Zhang composites appear smoother due to the weighted average within a 3 km search radius.

Scan type	26 Aug 2011	12 Apr 2012	06 Jun 2012	25 Aug 2012	All
RHI	95	106	69	86	356
PPI	309	215	279	297	1100

Note: For RHI scans, these numbers exclude the additional scans along the same azimuth as performed during the DYMECS study. For PPI scans, the number refers to scans at individual elevations.
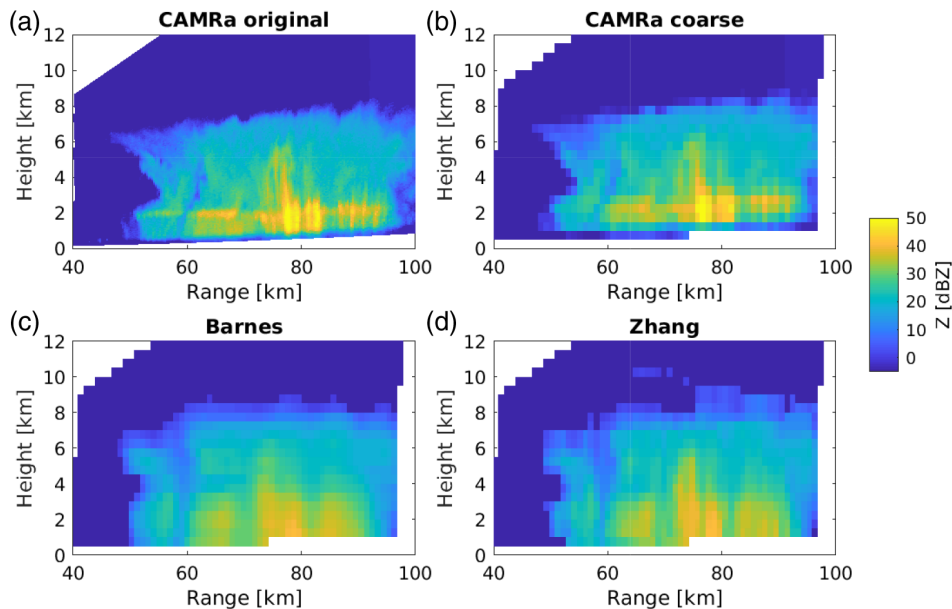
FIGURE 2 Example comparison between the interpolated CAMRa scan and the Barnes and Zhang composites. (a) CAMRa scan on the original grid (300 m range resolution, 0.048° elevation sampling). (b) CAMRa scan interpolated onto the 3D composite grid of 1 km horizontal and 500 m vertical resolution. (c) Barnes and (d) Zhang composites for the same (x,y) coordinates as (b)

We note that Lakshmanan *et al.* (2013) found empirically that for 18 dBZ ETH detection, interpolation in dBZ produced better results, which was confirmed by Warren and Protat (2019). However, Warren and Protat (2019) recommend interpolation in linear Z rather than dBZ in situations of high reflectivity and strong gradients, such as convective cores, which will be present in most of our scans. While we found some differences in our statistics depending on whether we interpolated the RHI scans in linear Z or dBZ units, these differences did not affect the interpretation of our results presented below.

If only a single range gate contributed to the interpolated value at x_i, the interpolated value was not considered in the subsequent analysis. The interpolated value was also not considered if one of the original CAMRa profiles measured $Z \geq 0$ dBZ at its greatest elevation, because ETH cannot be determined in that situation. Typically, this restriction was precluded by only considering RHI data within 40–100 km distance from Chilbolton, which was chosen to match the range considered with the PPI volume scans obtained with CAMRa. For each averaged profile, cloud layers were identified by requiring exceedance of either the 0 dBZ or the 10 dBZ threshold over a minimum vertical depth of 2 km. ETH was then determined as the maximum height of the top layer in the averaged profile.

The same ETH determination was performed on the radar composite data.

We also wish to establish the reliability of fractional cover in the radar composite data, since different radar reflectivity thresholds can be used to identify the extent or morphology of convective storms. In this context, fractional cover is the fraction of relevant data points (e.g., at a given height) above a threshold value. Both RHI and PPI scans from CAMRa are used to estimate fractional cover, with the RHI scans interpolated as described above for the ETH comparison. For the PPI scans, each scan was associated with the radar composite using the 5 min window ending after the time of the PPI scan (as with the RHIs). The CAMRa PPI scan was then linearly interpolated in dBZ units to the 1 km horizontal and 500 m vertical grid used in the composite and the range of observations was again restricted to 40–100 km distance from Chilbolton. For the composites, data points were only included in the analysis if the interpolated CAMRa scan observed the relevant point during the time window. As with the RHI comparison, the CAMRa and network data cannot be guaranteed to be collocated in space and time, and storm displacements between the two sets of data are assumed to be unbiased. At S-band, attenuation through rainfall is normally deemed negligible (Testud *et al.*, 2000).

Hail may lead to significant attenuation, but convective storms of such intensity are rarely observed across southern England. For instance, using the criteria for hail detection from Bringi *et al.* (1984), only 13 out of 356 RHI scans considered in this study were identified as potentially containing hail. Therefore, no attenuation correction was performed on the CAMRa data for the purpose of this study.

To estimate the uncertainty in ETH distribution and fractional cover from the CAMRa scans, we use a bootstrapping approach (e.g., Liu *et al.*, 2010). From a given population of N scans, a new population of N scans are sampled with replacement; for the RHIs, $N = 356$ (Table 1). This way, 1,000 new populations are formed, providing 1,000 estimates of the ETH distribution and fractional cover. From these 1,000 estimates, we will then report the median and interquartile range, as well as the 90% confidence interval, to establish how the radar composite statistics fall within the range of uncertainty of the limited CAMRa sample.

2.3 | Model simulations

The model evaluation presented in Section 4 focuses on a single day, 25 August 2012. The synoptic situation and evolution for this case is described in detail by Hanley *et al.* (2015) and has been studied for storm morphology, storm evolution, and updraught statistics (Stein *et al.*, 2014; Stein *et al.*, 2015; Nicol *et al.*, 2015). Model simulations were performed with the MetUM as part of these previous studies, and the same simulations are considered here, focusing only on the 1,500 and 200 m grid length simulations. Although the MetUM has since undergone various scientific changes, the primary purpose of this paper is to present a new model evaluation strategy. Furthermore, using the same simulations as Stein *et al.* (2014) allows for a direct comparison with the results presented previously.

The 1,500 m simulation was run with the same configuration as the operational forecast version of the MetUM at the time (version 7.8) and initialised at 0400 UTC; Hanley *et al.* (2015) provide details regarding the model design, but the relevant items for this study follow below. At 1,500 m grid length, the model runs without a convection parametrization scheme and has 70 vertical levels, which increase in spacing quadratically with height. Sub-grid horizontal mixing is parametrized using a Smagorinsky–Lilly type of scheme and vertical mixing is treated using the Lock *et al.* (2000) non-local boundary-layer scheme. The MetUM uses a single-moment microphysics scheme, which considers ice, liquid, and rain as prognostic variables.

At 200 m grid length, the MetUM was run with the same microphysics parameters, but with double the number of vertical levels. The sub-grid mixing was treated by a Smagorinsky–Lilly type scheme in both the horizontal and the vertical directions, using a mixing-length parameter of 40 m. While Hanley *et al.* (2015) demonstrated that the model representation of convective storms is highly sensitive to this parameter, the value of 40 m corresponds to a ratio of 0.2 between mixing length and grid length, which has been shown to best resolve turbulent eddies (Mason, 1994). The 200 m simulation was one-way nested in a 500 m simulation with the MetUM and initialised at 0700 UTC; the 500 m simulation was one-way nested in the aforementioned 1,500 m simulation, both initialised at 0400 UTC.

A "forward model" was run offline to calculate radar reflectivity for each model grid box based on these cloud fields and the relevant microphysics parameter choices (Stein *et al.*, 2014); the radar reflectivity was available only as hourly output. No smoothing was applied to the model data to attempt to match the radar data. The effective resolution of an NWP model is typically several times greater than the grid length (Skamarock, 2004), so that even at 200 m grid length, the MetUM cannot be considered to resolve scales less than 1 km. Therefore, we only evaluate the model performance for features of sizes that can be expected to be resolved by the 200 m grid-length simulation and observed within the radar composite data. The minimum size of the convective storms considered is based on a surface rainfall rate greater than $4 \, \mathrm{mm \cdot hr^{-1}}$ over an area of at least $4 \, \mathrm{km^2}$.

3 | EVALUATION AGAINST CAMRA

3.1 | Cloud-top height

The ETH distributions from the collocated RHI and composite data are shown in Figure 3. We see that both the Zhang and Barnes techniques produce similar ETH distributions to CAMRa, although the composites retrieve too many ETH_0 greater than 8 km and too few ETH_{10} and ETH_{18} greater than 8 km. The cumulative distribution functions (CDFs) shown in Figure 3d–f, show that CAMRa has a 99%ile of ETH_0 approximately 9.5 km, while the Zhang and Barnes composites have 99%iles of 10.5 and 10 km, respectively. For ETH_{10} and ETH_{18}, the Zhang composite matches the CAMRa distribution well, whereas the Barnes composite tends to underestimate ETH by about 0.5 km between the 90 and 99%ile. As a result of these differences, fractional cover of clouds above a given ETH threshold can vary drastically between the composites and
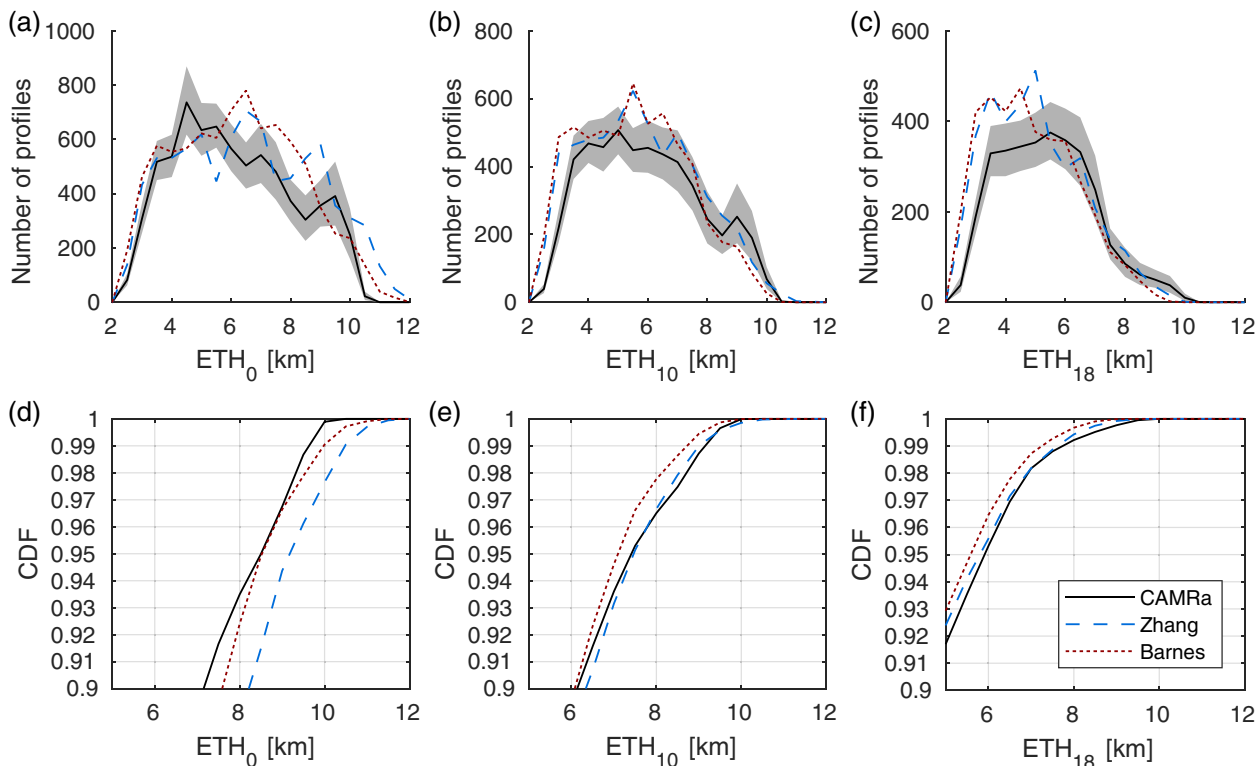
FIGURE 3 Cloud-top height statistics derived from RHI scans using (a, d) a 0 dBZ threshold, (b, e) a 10 dBZ threshold, and (c, f) an 18 dBZ threshold. (a–c) show distributions of cloud-top heights for the interpolated CAMRa data and near-coincident composites using the Zhang (blue dashed) and Barnes (red dotted) algorithms. Grey shading for the CAMRa distribution indicates the bootstrapped 90% confidence interval. (d–f) Cumulative distribution function (CDF) for each cloud-top height bin, for the different observations

CAMRa. A 9 km threshold at ETH_0 could lead to twice as much cover estimated from the Zhang composite than with CAMRa (or the Barnes composite). An 8 km threshold at ETH_{10} could lead to only half as much cover estimated from the Barnes composite compared to CAMRa (or the Zhang composite).

When we compare ETH retrievals for individual columns in Figure 4, both the Barnes and Zhang composites produce medians and interquartile ranges of ETH that compare well against particular CAMRa ETH observations for ETH_0 and for $ETH_{10} \leq 8$ km. Looking specifically at high ETH values, when $ETH_0 > 8$ km is detected with CAMRa, the Zhang composite tends to overestimate ETH_0 (Figure 4a), while the Barnes composite has a narrow interquartile range for these profiles (Figure 4d). This finding is consistent with the CDF of Figure 3, suggesting that the Barnes composite may be preferred for the purpose of ETH_0 identification. For the ETH_{10} and ETH_{18} thresholds, both composites underestimate ETH for the deepest profiles observed by CAMRa, with particularly Barnes frequently missing 18 dBZ layers (Figure 4f). In Table 2, we report the bias and RMSE of the Barnes and Zhang composites compared to CAMRa for ETH_0, ETH_{10}, and ETH_{18}. The Barnes composite has a lower RMSE and absolute bias than Zhang when all profiles are

considered. We suspect that the low bias at higher thresholds is due to the many correct negatives, that is, both CAMRa and the composite detect no cloud, or ETH=0. When we restrict our analysis to profiles where a 2 km thick layer of the relevant reflectivity is identified in both CAMRa and the composite, the Barnes composite has a better RMSE for all thresholds, but the Zhang composite has a smaller absolute bias with the 10 and 18 dBZ thresholds. The result that the Barnes composite performs better at a 0 dBZ threshold and the Zhang composite at the 10 and 18 dBZ thresholds agrees with the ETH distribution findings in Figure 3. The correlations do not improve when we impose the condition that both composite and CAMRa detect a cloud layer. This may be due to many correct negatives (none detect a cloud layer) balancing out the false alarms and misses when we do not impose this condition.

This lack of improvement in correlation as well as the large spread indicated by the 90% confidence interval in Figure 4 are likely due to inaccuracies when the data are collocated in space and time. We therefore consider the detection statistics, that is, the frequency of a cloud layer being detected in each of the composites and in the CAMRa data. From these statistics in Tables 3 and 4, we note that 35.8% of the CAMRa averaged profiles contain
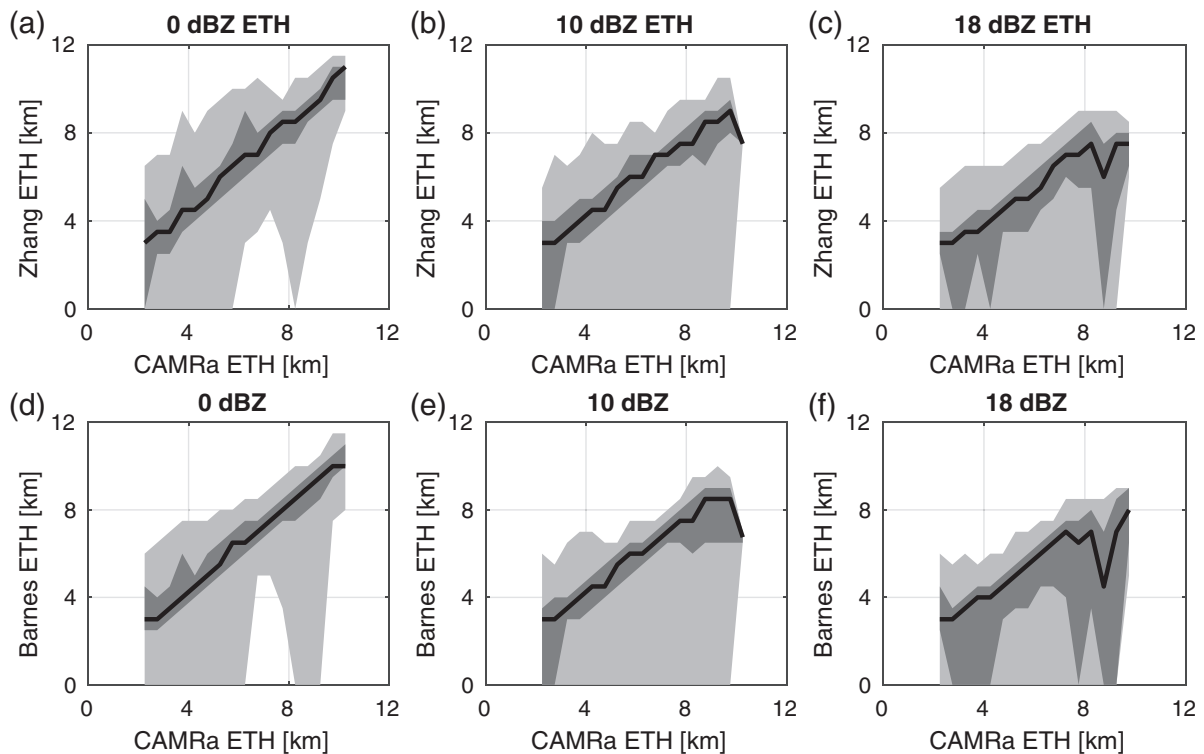
FIGURE 4 Direct comparison of cloud-top height (ETH) derived using the CAMRa averaged data and (a–c) the Zhang composite, and (d–f) the Barnes composite. (a, d) use a 0 dBZ threshold to determine ETH, (b, e) a 10 dBZ threshold, and (c, f) 18 dBZ. For each CAMRa 500 m ETH bin, the interquartile range (dark grey), median (black) and 90% confidence interval (light grey) are shown

TABLE 2 Error statistics in ETH estimates against CAMRa (RMSE and bias in km)

Threshold	Barnes			Zhang		
	RMSE	Bias	Correlation	RMSE	Bias	Correlation
0 dBZ	2.11	0.41	0.80	2.27	0.52	0.79
10 dBZ	1.87	0.12	0.77	1.93	0.20	0.76
18 dBZ	1.62	0.03	0.68	1.66	0.09	0.68
0 dBZ detected	1.28	0.30	0.81	1.64	0.59	0.74
10 dBZ detected	1.24	−0.15	0.76	1.33	0.03	0.73
18 dBZ detected	1.20	−0.27	0.69	1.27	−0.16	0.67

Note: The first three rows are for all profiles within 40–100 km range of Chilbolton. The lower three rows are for profiles in the same range, but conditional on both CAMRa and the composite identifying a 2 km-thick cloud layer of the specified radar reflectivity.

a 0 dBZ cloud layer with thickness greater than 2 km, but the Barnes composite misses cloud (above 0 dBZ) in 9.5% of those profiles and Zhang 10.1%. Based on examples such as Figure 2, we assume that the "misses" are mostly due to upper-level clouds such as anvils extending from the convective storms, which are poorly sampled by the UK radars. We note for instance that the highest-elevation scan of the network radars is typically 4°, which reaches 8 km just beyond a range of 100 km, where the beam will

be approximately 1.75 km wide. When we condition the statistics on CAMRa detecting a 10 dBZ or 18 dBZ layer, Barnes detects no cloud (above 0 dBZ) in only 4.3% and 2.5% of these cases, respectively, and Zhang in 5.1% and 2.5%. This improvement in cloud detection supports the suggestion that upper-level cloud is missed, as these clouds tend to be associated with lower reflectivity values.

The Barnes and Zhang composites detect cloud in 41.9% and 41.5% of the profiles, respectively, but no cloud

TABLE 3 Detection statistics of the Zhang composite in per cent

	Zhang no cloud	Zhang ETH_0	Zhang ETH_{10}	Zhang ETH_{18}
CAMRa no cloud	54.9	9.3	3.9	1.6
CAMRa ETH_0	3.6	32.2	25.8	17.6
CAMRa ETH_{10}	1.3	24.2	21.8	15.9
CAMRa ETH_{18}	0.4	15.7	14.8	12.5

Note: The sum of the first two elements in the first row is the total frequency of CAMRa detecting no cloud, and the sum of the first two elements in the second row is the total frequency of CAMRa detecting cloud, as "no cloud" implies that no 0 dBZ layer of 2 km thickness was detected. Added together, these two sums equal 100. Note that the frequency of ETH_{18} is included in that of ETH_{10} and that the frequency of ETH_{10} is included in ETH_0.

TABLE 4 As Table 3, but for the Barnes composite

	Barnes no cloud	Barnes ETH_0	Barnes ETH_{10}	Barnes ETH_{18}
CAMRa no cloud	54.7	9.5	3.6	1.4
CAMRa ETH_0	3.4	32.4	25.8	17.2
CAMRa ETH_{10}	1.1	24.4	21.7	15.6
CAMRa ETH_{18}	.4	15.7	14.7	12.3

(above 0 dBZ) is detected in the CAMRa data in 29.3% and 22.4% of those profiles. Again based on examples such as Figure 2, we assume that these "false alarms" are due to smearing out of the convective storms during the processing of the composites, especially due to the search radius. Considering 10 and 18 dBZ layers, CAMRa does not detect cloud (above 0 dBZ) when these are identified in the Barnes composite in 12.2% and 7.5% of the cases, and in 13.1 and 8.3% of these layers identified in the Zhang composite.

These comparisons suggest that ETH identification using the Barnes and Zhang composites should not be used to study *individual* columns in the composites. The use of an ETH threshold to study coverage of high-topped clouds should be done with caution as, for different choices of composite and reflectivity, the coverage can be over- or underestimated. However, when considering a set of ETH values for a given time or location, for instance associated with a cloud or storm object such as that shown in Figure 2, the Barnes and Zhang composites can be relied upon to determine the mean ETH. For statistics calculated over many profiles of many clouds, as illustrated in Table 2 and Figure 4, the Barnes and Zhang composites compare very well against the CAMRa profiles.
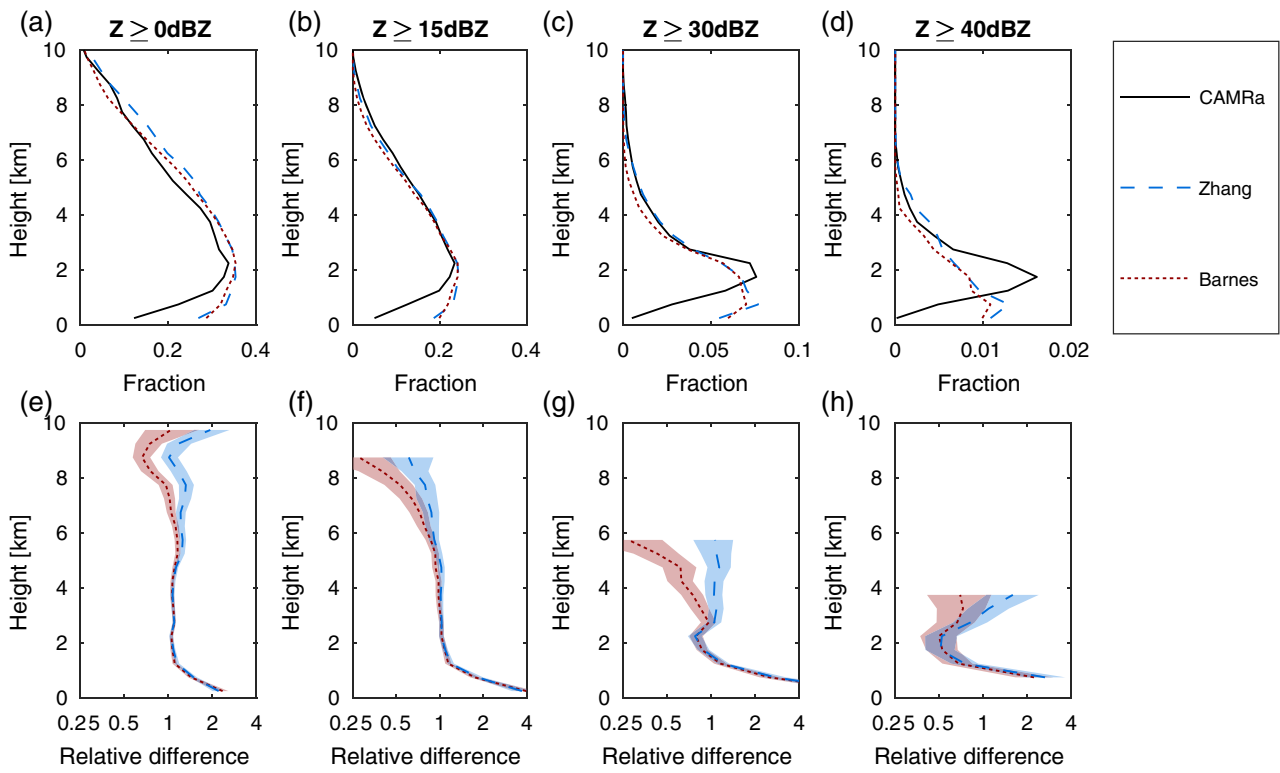
3.2 | Fractional cover

The purpose of this paper is to verify whether the radar composites can be reliably used to study the morphology of deep convective storms, including their internal macrophysical structure such as the size of cores of intense radar reflectivity. Before analysing storm morphology statistics, we wish to ascertain that the radar composite retrieves fractional cover for different radar reflectivity thresholds comparable to the CAMRa data. Such analysis should assist in the interpretation of differences in average storm structures in the different datasets.

Figures 5 and 6 show the fractional coverage in the different datasets for four radar reflectivity thresholds in the RHI scans and PPI sector scans, respectively. The differences between the two figures can be attributed to the scan strategy: RHI scans targeted the locations of most intense rainfall and may therefore be expected to have greater occurrence of $Z \geq 30$ dBZ overall and of $Z \geq 0$ dBZ at higher altitudes. The Zhang composite consistently retrieves more cover than the Barnes composite for all thresholds considered. This leads to a better performance by Zhang for the high thresholds, but a slightly larger overestimate of fractional occurrence for $Z \geq 0$ dBZ above 5 km compared to CAMRa, whereas the Barnes composite performs better for this lowest threshold. The Zhang composite also retrieves $Z \geq 0$ dBZ above 9 km too often, which is consistent with the finding that it overestimates ETH_0.

Both the Zhang and the Barnes composites significantly underestimate the occurrence of $Z \geq 30$ dBZ and $Z \geq 40$ dBZ between 1 and 3 km, although the Zhang fractional cover matches the CAMRa cover very well above 3 km. Inspection of individual RHI scans indicates that many storms were surrounded by substantial stratiform regions, for which a radar bright band was present (e.g., Figure 2). Since the bright band is typically only a few hundred metres deep, it will often be missed by the UK radars due to their scan strategy and therefore will not be prominent in the 3D composite. It is likely that the bright band appears in the CAMRa PPI scans because, typically, 10–13 scans were performed to capture the full height of convective storms, allowing the bright band to be sampled multiple times in individual storms.

It is not immediately clear why the Zhang composite retrieves higher fractional cover than the Barnes composite. The Zhang method finds the nearest neighbour and then applies vertical interpolation. Conversely, for the Barnes composite, observations below the noise value are given a value of −10 dBZ in order to ensure that the iterative technique behaves sensibly near the rain/no-rain boundary. Once the final (smoothed) field is obtained, the field is then thresholded again, at 0 dBZ, to retrieve the output. The extent to which the coverage in the Barnes composite

FIGURE 5 Fractional coverage for different radar reflectivity factors based on RHI scans for CAMRa (black), Barnes (red), and Zhang (blue dashed). Thresholds are: (a, e) $Z \geq 0$ dBZ, (b, f) $Z \geq 15$ dBZ, (c, g) $Z \geq 30$ dBZ, and (d, h) $Z \geq 40$, dBZ. (e–h) show the bootstrapped median and 90% confidence interval for the relative difference in fractional coverage with CAMRa for Zhang (blue dashed and shading) and Barnes (red dotted and shading). Only heights for which all three fractions are greater than 0.1% are shown
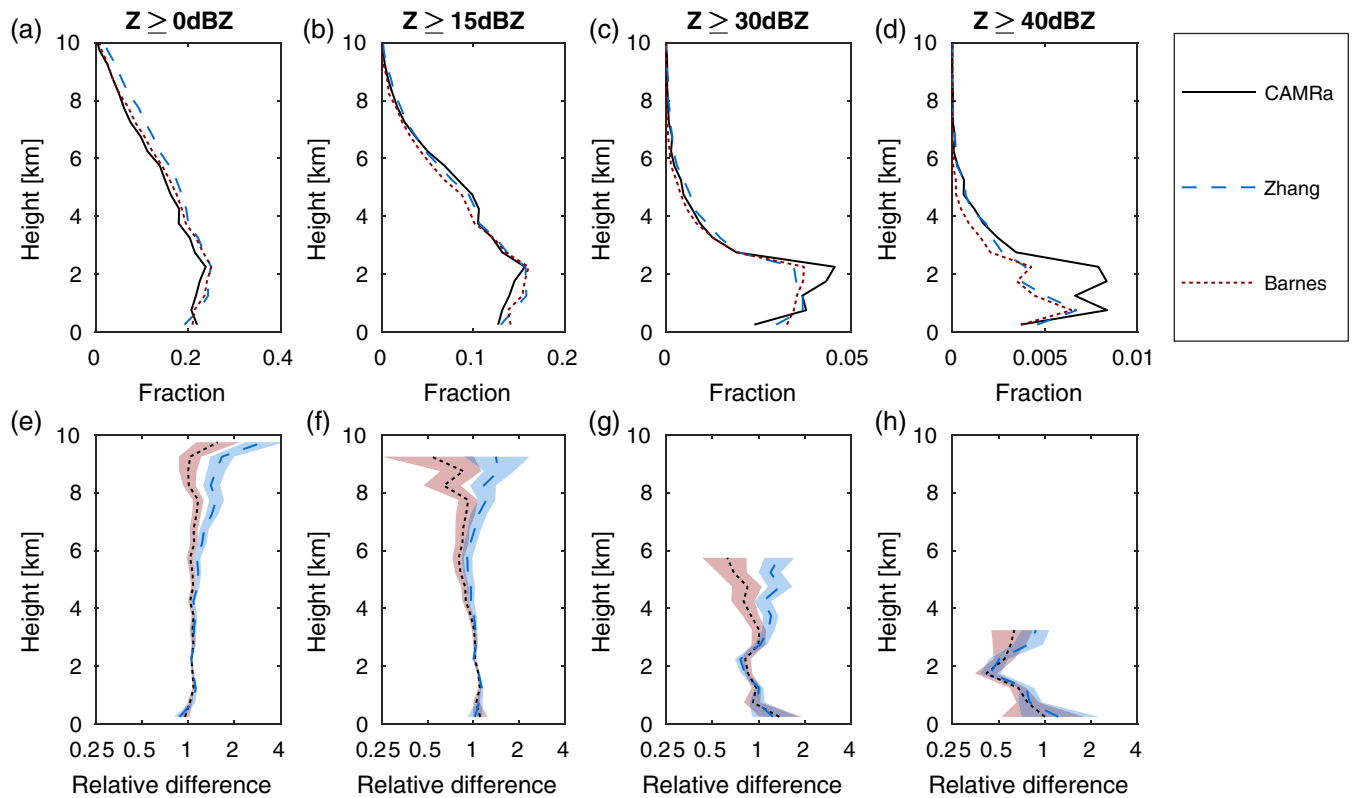

FIGURE 6 As Figure 5, but based on PPI scans

is affected by the horizontal interpolation will depend on the amount of no-rain in the area, the chosen smoothing widths, and the choice of threshold.

The lack of high radar reflectivity retrievals between 1 and 3 km becomes an issue if radar reflectivity thresholds are used at low altitudes to identify individual storm features. Potentially, the maximum radar reflectivity in the column may be used instead. In Figure 7, the distribution of maximum radar reflectivity is shown for the CAMRa data, Barnes and Zhang composites, restricted to the CAMRa RHI scans as in the previous subsection. The Barnes composite traces the CDF of CAMRa very well up to 25–30 dBZ, suggesting that fractional cover up to this threshold is represented well by the Barnes composite. The Zhang composite tends to have slightly lower CDF values than CAMRa up to 30 dBZ, suggesting that its fractional cover might be slightly overestimated for lower reflectivity thresholds. At higher reflectivity values, the Zhang composite compares well against CAMRa, whereas the Barnes composite underestimates the reflectivity values associated with high percentiles. In particular, the 95 and 99%iles for CAMRa are approximately 33.6 and 42.2 dBZ, compared to 33.5 and 41.8 dBZ in the Zhang composite and 32.8 dBZ and 40.4 dBZ in the Barnes composite. Conversely, for thresholds of 30 and 40 dBZ, the CAMRa coverage is 8% and 1.5%, the Zhang composite coverage is 7.9% and 1.3%, and the Barnes composite coverage is 7.4% and 1.1%. This comparison implies that the radar composites will underestimate storm sizes considerably if storms are identified with a threshold of $Z_{max} \geq 40$ dBZ and therefore a lower threshold, ideally no higher than 30 dBZ, is advised. Also, based on these results and those in Figures 5 and 6, it is advised to use Z_{max} for storm detection in the radar composite rather than Z at a fixed low altitude.

3.3 | Storm morphology

From the CAMRa observations during DYMECS, Stein *et al.* (2014) concluded that the Met Office 1,500 m model simulated deep convective storms that were a factor 1.5–2 larger than observed. This was based on storms observed throughout the day as, due to the nature of the scanning strategy, no distinction between storm sizes at different times of the day was possible. The driving purpose of the present study is to determine whether the 3D radar composites can be used to study the evolution of convective storms – that is, the storm height and width, as well as intensity – throughout the day and to evaluate these characteristics in NWP model simulations. We therefore wish to assess if the radar composites compare well against the CAMRa data in terms of storm morphology.
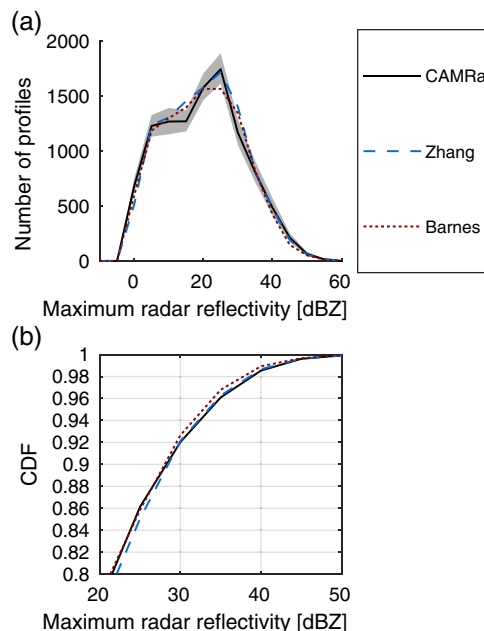
FIGURE 7 Distribution of maximum radar reflectivity in a column derived from RHI scans for ranges from Chilbolton between 40 and 100 km. Profiles were not included if the CAMRa radar reflectivity at the highest elevation was greater than 0 dBZ. (a) Distribution of maximum Z for CAMRa and near-coincident composites using the Zhang and Barnes algorithms; bins are 5 dB. Grey shading indicates the bootstrapped 90% confidence interval. Clear profiles for CAMRa have values below 0 dBZ, whereas clear profiles for the composites have values set to –999. (b) CDF at each 5 dB radar reflectivity bin

Individual storms are identified from the PPI sector scans with CAMRa following the methodology described in Stein *et al.* (2014). For each PPI scan in a set, all observation points are advected to a uniform time (typically to the nearest 5 min time stamp) using a single storm motion vector. This advection vector is derived from the radar composite data using the spatial correlation between 5 min rainfall fields. The underlying assumptions are that

(a) a single storm motion vector is representative for all heights,

(b) the storm motion vector is fixed during the scanning period, and

(c) effects of storm evolution on morphology are negligible within this time frame.

Note that this approach is the same as the advection correction applied to generate the radar composite, with the latter using optical flow on the vertically integrated liquid from the two previous composites, rather than correlation on the rainfall product derived from the two previous composites. In the present study, a 3D reconstruction of an individual storm is then realised by considering at each height all regions (identified using 8-connectivity) with $Z \geq 0$ dBZ that overlap with the storm volume identified
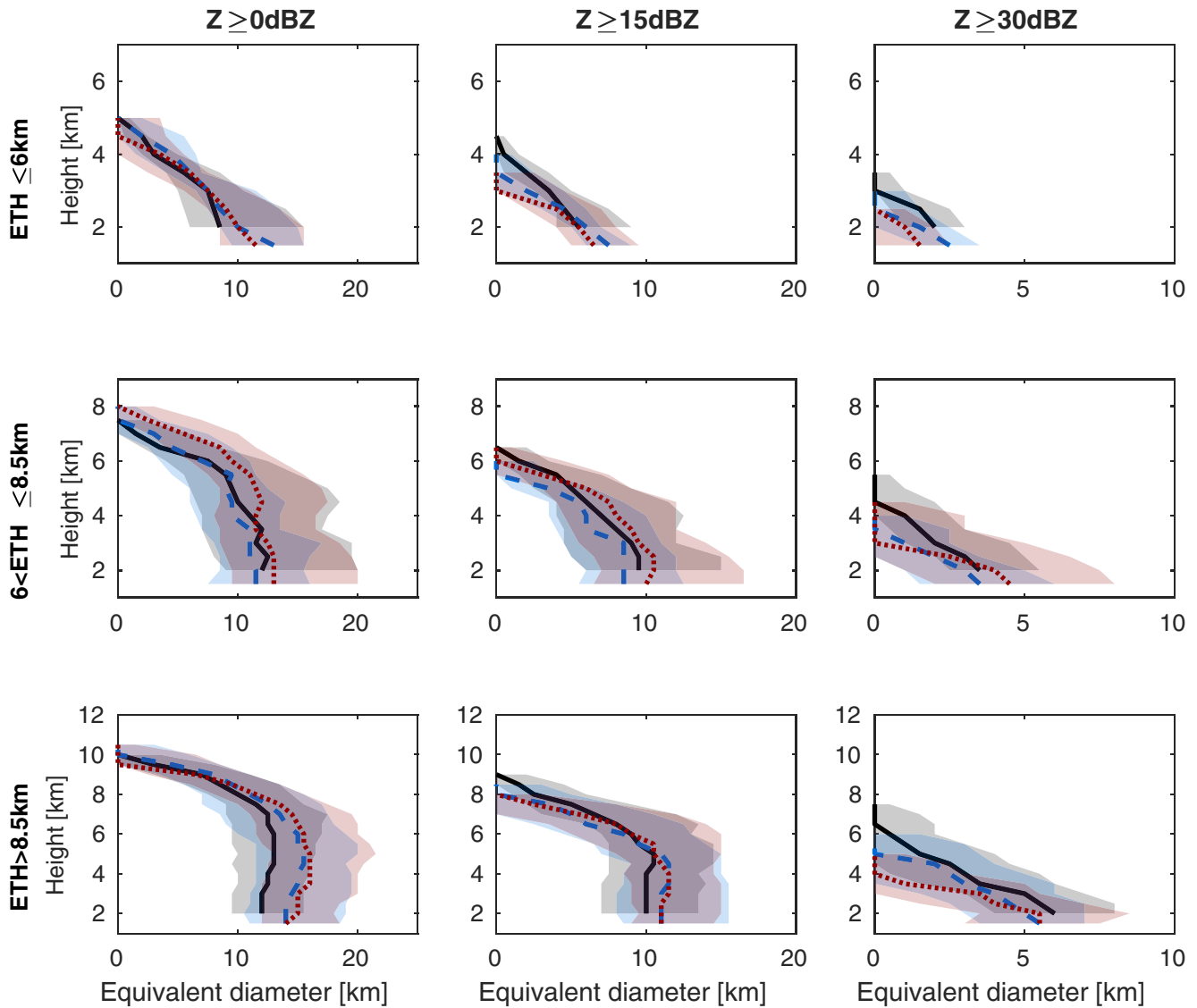
FIGURE 8 Median (solid lines) and interquartile range (shading) of equivalent diameter of storms observed by CAMRa across the four days, classified by ETH_0 identified by CAMRa as being (a–c) up to 6 km, (d–f) 6–8.5 km, and (g–i) > 8.5 km. Equivalent diameters are shown for radar reflectivity thresholds of (a"d, g) 0 dBZ, (b, e, h) 15 dBZ, and (c, f, i) 30 dBZ. The CAMRa results are shown in black and solid, Zhang in blue and dashed in red and dotted. The CAMRa results are cut off at 2 km, as the lowest elevation scanned will typically not observe below this height throughout the domain

thus far. This process starts at 1 km above mean sea level (which is above ground for southern England), considering the region with $Z \geq 0$ dBZ that encompasses the storm feature originally identified using $Z_{max} \geq 32$ dBZ.

For the comparison presented here, when a storm is identified in a CAMRa volume scan, its location is matched to the 3D radar composite data for the same time stamp and a storm is reconstructed in the 3D radar composite data (from the lowest level upwards) similar to the method used for the CAMRa data. In Figure 8, the median equivalent diameter is shown for storms classified by their maximum ETH_0 and for different radar reflectivity thresholds.

For each storm at a given height, the equivalent diameter is recorded as the diameter of a circle with the same area as the total area of the storm with $Z \geq Z_t$ at that height. The reduction in storm size below 2 km for CAMRa that is evident in Figure 8 is due to the reduction of radar observations close to the surface, as these will typically be below the lowest elevation scanned. This reduction was accounted for by Stein *et al.* (2014) by projecting the lowest CAMRa PPI scan to the surface, which was not done for the present comparison. Also, note that due to differences in ETH_0 detection, storm numbers in each category differ slightly between the three datasets (Table 5).

TABLE 5 Total number of storms identified per ETH_0 class in the CAMRa data and the two radar composites

Data	ETH_0 ≤ 6 km	$6 < ETH_0$ ≤ 8.5 km	ETH_0 > 8.5 km
CAMRa	23	40	67
Barnes	28	45	57
Zhang	21	27	82

The storm morphology results are consistent with those for ETH and fractional cover. For the 0 dBZ threshold, storms tend to be slightly larger in the radar composites than observed with CAMRa, although the difference in median is at most 25%, much less than the difference between the Met Office 1,500 m model and the CAMRa observations reported by Stein *et al.* (2014), and typically within the interquartile range of the CAMRa observations. For the 15 dBZ threshold, the radar composites compare very well against CAMRa, consistent with the good agreement of fractional cover for $Z_{max} \geq$ 15 dBZ. For the 30 dBZ threshold, the radar composites tend to lack observations at these high radar reflectivity values as noted in the previous subsection, and as a result the equivalent diameter is less than observed by CAMRa.

Together with the analysis of ETH and fractional cover, these results support the use of the 3D radar composites for model evaluation of convective storm characteristics. However, all results indicate that the 3D composites lack retrievals of high radar reflectivity and that therefore the use of thresholds above 15 dBZ will lead to underestimates of fractional cover and storm sizes. It is unlikely this is due to the raw radar measurements lacking these high reflectivity observations, as illustrated by our analysis of Z_{max} in Figure 7. For instance, Scovell and Al-Sakka (2016) presented a case-study of a supercell thunderstorm with reflectivity values exceeding 50 dBZ in a core about 10 km across, and such high values are present in the composite too. However, in the DYMECS cases studied here, many cores were at most a few km across (e.g., Nicol *et al.*, 2015). We hypothesise that both the Zhang and Barnes composites can often underestimate the peak reflectivity values for small convective cores because the peak dBZ measurements are diluted by the lower dBZ ones, due to a combination of the radar-to-Cartesian grid interpolation and possible inaccuracies in the temporal matching of individual scans. Such inaccuracies are less likely to affect statistics for a larger convective core, such as the one reported above, as there will be more cases where the pixel averages contain data exclusively from values near the peak reflectivity.
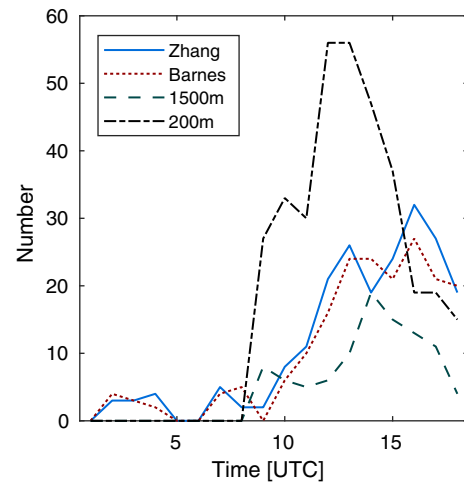
FIGURE 9 Total number of storms in each of the datasets against time of day

4 | MODEL EVALUATION

In this Section, we present results of a model comparison against the Zhang composite. Our primary interest is to expand on results presented in Stein *et al.* (2014), namely by considering storm structures at different times of the day. The results in the previous sections indicate that

(a) composite statistics from a large number of storms show that the Zhang and Barnes composites lead to similar storm structures as the CAMRa observations used by Stein *et al.* (2014), and

(b) thresholds above 15 dBZ will lead to underestimates in fractional cover compared to the CAMRa observations.

In the following model evaluation, we therefore restrict ourselves to the 0 dBZ threshold only.

The total number of storms considered is shown in Figure 9. Both the 1,500 and 200 m MetUM have initiated convection between 0800 and 0900 UTC, with storm numbers sharply increasing, whereas in the observations the storm numbers gradually increase from 0900 to 1300 UTC. At 200 m grid length, the model generates twice as many storms as observed, whereas at 1,500 m grid length, the model produces fewer storms than observed. Hanley *et al.* (2015) showed that the 1,500 m MetUM also had a lower domain-averaged rainfall rate for this case, whereas the 200 m simulation compared well against the radar-derived rainfall rates until 1300 UTC. Furthermore, in the observations, the domain-averaged rainfall rate nearly doubled between 1300 and 1500 UTC, which is in contrast with a near-constant number of storms; this warrants further investigation. Regarding the model performance, Hanley *et al.* (2015) found that the representation of storm sizes and life cycles strongly depended on the mixing length chosen in the sub-grid turbulence parametrization scheme.

FIGURE 10 Hourly statistics of storm morphology on 25 August 2012 for the Zhang composite, 1,500 m MetUM simulations, and 200 m MetUM simulations. For each hour (columns, times in UTC), the median (light colours), 25th (dark) and 75th percentiles (white) of equivalent radius are shown as "distance from centre". The percentiles are taken over all storms identified on the hour from features with surface rainfall rates > 4 mm·hr^{-1} over an area of at least 4 km^2. The radius is mirrored across the y-axis at each time for ease of comparison of storm scales across datasets and across times
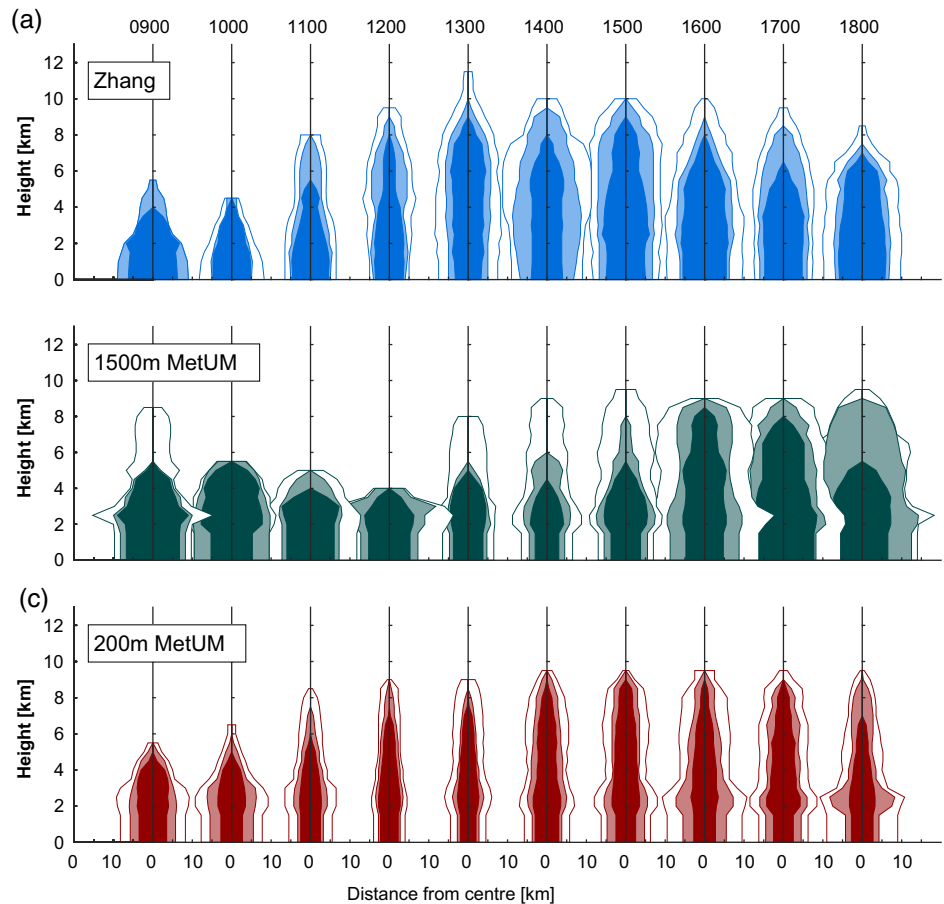
In Figure 10, we show the median and interquartile range of storm sizes at each hour for the Zhang composite, the 1,500 and 200 m MetUM simulations. The majority of identified storms in the Zhang composite have ETH below 8 km until some time between 1100 and 1200 UTC. Between 1200 and 1600 UTC, less than 25% of storms in the Zhang composite have ETH below 8 km. Storms also grow wider throughout the day, with the 25%ile at its maximum width of 22 km at 1600 UTC. The MetUM at 1,500 m grid length shows a vastly different cycle of storm morphology than the Zhang composite. This is not surprising since, at this grid length, convective storms (and even more so the updraughts within the storms) will be under-resolved (Nicol *et al.*, 2015). A majority of ETH greater than 8 km is not reached until 1500 UTC, whereas the high cloud tops dominate until 1700 UTC, later than in the Zhang composite. Storms do grow wider throughout the day starting from 1300 UTC. Interestingly, the median storm width compares very well with the Zhang composite between 1300 and 1500 UTC, although storms are generally shallower than observed. At 200 m grid length, the MetUM shows broadly similar behaviour to the Zhang composite. The 8 km ETH is reached by the majority of storms at 1200 UTC and more than 75% have this ETH or greater from 1300 to 1700 UTC. However, at this grid

length, storms are generally less wide than observed, as the 75%ile is generally smaller than the observed median diameter.

Further analysis of this simulation is out of scope of this paper, and the MetUM scientific configuration has advanced since this simulation was originally carried out. However, the purpose here is to demonstrate the type of model biases that can be identified using the 3D radar composite.

5 | CONCLUSION

We have shown that a 3D radar composite over the UK can reliably be used to study certain convective storm characteristics. The availability of the operational radar data at 5 min frequency allows for new analysis regarding the timing and evolution of convective storms (or any precipitating clouds), and the routine evaluation of such characteristics in NWP models. The archive of the radar data extends back to July 2010, enabling the development of new understanding of typical convective storm behaviour in different synoptic conditions, for instance using composite statistics, which is not possible using sparsely available data from targeted field campaigns.

Two compositing methods were analysed, namely one derived using the Barnes (1964) technique and one derived using the Zhang *et al.* (2005) technique. For four cases of widespread convective storms, both 3D radar composites had similar fractional cover with height at radar reflectivity thresholds of 0 and 15 dBZ compared to high-resolution radar observations from CAMRa. At radar reflectivity thresholds of 30 and 40 dBZ, the fractional cover was underestimated, particularly at lower elevations. Both 3D radar composites led to similar distributions of ETH compared to the CAMRa observations, although slightly higher ETH were observed at a 0 dBZ threshold. When the composites were compared to CAMRa at the grid scale, the Barnes (1964) technique had lower bias and RMSE, and higher correlation at all thresholds. Both composite techniques showed a large spread in ETH values for a given CAMRa observation, although this could partly be due to difficulties with spatio-temporal collocation of the two datasets. Storm morphologies were obtained from the 3D composites and for collocated sector volume scans with CAMRa. For these selected storms, for thresholds of 0 and 15 dBZ, equivalent storm diameter with height derived from the 3D composite compared well against the profiles derived from CAMRa, indicating that the analysis from Stein *et al.* (2014) can be reproduced using the 3D composites. At a 30 dBZ threshold, the morphology derived from the 3D composites did not stretch high enough, likely due to smoothing out of convective cores.

We demonstrated the potential use of the 3D radar composite for the evaluation of convective storm characteristics in convection-permitting simulations. The analysis of storm morphology confirmed the results from Stein *et al.* (2014), namely that at 1,500 m grid length the MetUM produced storms that were too wide, and at 200 m grid length storms were too narrow. However, the 5 min frequency of the 3D radar composite allowed the extension of this analysis to study the diurnal cycle of storm characteristics, highlighting potential issues with convective storm initiation, growth, development, and organisation for both model grid lengths.

The results presented show promising potential applications of the 3D radar composites, but further development of the techniques should be considered. Regarding the ETH, Lakshmanan *et al.* (2013) and Warren and Protat (2019) indicate that interpolation in dBZ will give better results. However, Warren and Protat (2019) suggest that, particularly for convective situations, interpolation in linear Z is recommended. A new 3D radar composite could consider a first step to interpolate in dBZ between scans in the vertical direction only, followed by the existing procedure to interpolate in 3D. A second extension of the existing procedure would be to include differential reflectivity, Z_{dr}, which could improve the detection of convective cores from columns of positive Z_{dr} (e.g., Kumjian *et al.*, 2014), which have previously been detected in UK convective storms (Plummer *et al.*, 2018).

Finally, it is worth considering the requirements on a radar network to reproduce this work with composites elsewhere. The German radar network has a similar dense spacing to the UK (e.g., Bick *et al.*, 2016) and in Europe more broadly, neighbouring radars are spaced within 128 km (Huuskonen *et al.*, 2014). In the eastern US, the typical spacing between NEXRAD radars is 235 km (Chen and Chandrasekar, 2018) and radars from the South African Weather Service are spaced a similar distance (Stein *et al.*, 2019). Given that both compositing techniques employ a multi-radar weighted average which depends on distance from each radar, individual points in the Cartesian mosaic will likely be dominated by observations from a single radar in these networks. Nevertheless, Zhang *et al.* (2005) developed their technique so that it could be used for a single radar as well as for a multi-radar mosaic. In particular, Feng *et al.* (2011) successfully used this 3D compositing technique on the NEXRAD radar data to classify various cloud layers within the 3D morphology of mesoscale convective systems. A future study comparing a 3D composite based on a single radar against CAMRa observations would be required to test the validity of such a composite, but the UK operational radars scan only at five elevations up to 5.0°. A dedicated campaign using the Wardon Hill radar, situated near Chilbolton and available to try different volume scan patterns, could enable this work to progress. Alternatively, comparisons of single-radar 3D mosaics against other research-grade and steerable radars would also require dedicated campaigns. However, such analysis and evaluation could strongly benefit future NWP model evaluation studies, in particular given the steady progression towards global convection-permitting forecasting systems (Bauer *et al.*, 2015).

ORCID
Thorwald H. M. Stein https://orcid.org/0000-0002-9215-5397

Robert W. Scovell https://orcid.org/0000-0002-9664-549X

Kirsty E. Hanley https://orcid.org/0000-0002-1002-7667

Humphrey W. Lean https://orcid.org/0000-0002-1274-4619

Nicola H. Marsden https://orcid.org/0000-0001-5985-7031

REFERENCES

Barnes, S.L. (1964) A technique for maximizing details in numerical weather map analysis. *Journal of Applied Meteorology*, 3, 396–409.

Bauer, P., Thorpe, A.J. and Brunet, G. (2015) The quiet revolution of numerical weather prediction. *Nature*, 525, 47–55. https://doi.org/10.1038/nature14956

Bick, T., Simmer, C., Trömel, S., Wapler, K., Hendricks Franssen, H.-J., Stephan, K., Blahak, U., Schraff, C., Reich, H., Zeng, Y. and Potthast, R. (2016) Assimilation of 3D radar reflectivities with an ensemble Kalman filter on the convective scale. *Quarterly Journal of the Royal Meteorological Society*, 142, 1490–1504.

Bringi, V., Seliga, T.A. and Aydin, K. (1984) Hail detection with a differential reflectivity radar. *Science*, 225, 1145–1147.

Chen, H. and Chandrasekar, V. (2018). Real-time wind velocity retrieval in the precipitation system using high-resolution operational multi-radar network, Chapter 15 in Remote Sensing of Aerosols, Clouds, and Precipitation. Islam, T., Hu, Y., Kokhanovsky, A., Wang, J. (eds). Elsevier, Amsterdam, Netherlands.

Clark, P.A., Roberts, N., Lean, H.W., Ballard, S.P. and Charlton-Perez, C. (2016) Convection-permitting models: a step-change in rainfall forecasting. *Meteorological Applications*, 23, 165–181. https://doi.org/10.1002/met.1538

Dance, S.L., Ballard, S.P., Bannister, R.N., Clark, P.A., Cloke, H.L., Darlington, T., Flack, D.L.A., Gray, S.L., Hawkness-Smith, L., Husnoo, N., Illingworth, A.J., Kelly, G.A., Lean, H.W., Li, D., Nichols, N.K., Nicol, J.C., Oxley, A., Plant, R.S., Roberts, N.M., Roulstone, I., Simonin, D., Thompson, R.J and Waller, J.A. (2019) Improvements in forecasting intense rainfall: results from the FRANC (Forecasting Rainfall exploiting new data Assimilation techniques and Novel observations of Convection) project. *Atmosphere*, 10. https://doi.org/10.3390/atmos10030125

Davis, C., Brown, B. and Bullock, R. (2006) Object-based verification of precipitation forecasts. Part 1: methodology and application to mesoscale rain areas. *Monthly Weather Review*, 134, 1772–1784.

Fabry, F. (2015) *Radar Meteorology: Principles and Practice*. Cambridge University Press, Cambridge, UK.

Feng, Z., Dong, X., Xi, B., Schumacher, C., Minnis, P. and Khaiyer, M. (2011) Top-of-atmosphere radiation budget of convective core/stratiform rain and anvil clouds from deep convective systems. *Journal of Geophysical Research: Atmospheres*, 116(D23). https://doi.org/10.1029/2011JD016451

Goddard, J., Eastment, J.D. and Thurai, M. (1994) The Chilbolton advanced meteorological radar: a tool for multidisciplinary atmospheric research. *Electronics and Communication Engineering Journal*, 6, 77–86.

Hanley, K.E., Plant, R.S., Stein, T.H.M., Hogan, R.J., Nicol, J.C., Lean, H.W., Halliwell, C.E. and Clark, P.A. (2015) Mixing-length controls on high-resolution simulations of convective storms. *Quarterly Journal of the Royal Meteorological Society*, 141, 272–284.

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

Hitschfeld, W. and Bordan, J. (1954) Errors inherent in the radar measurement of rainfall at attenuating wavelengths. *Journal of Meteorology*, 11, 58–67.

Huuskonen, A., Saltikoff, E. and Holleman, I. (2014) The operational weather radar network in Europe. *Bulletin of the American Meteorological Society*, 95, 897–907.

Kumjian, M.R., Khain, A.P., Benmoshe, N., Ilotoviz, E., Ryzhkov, A.V. and Phillips, V.T.J. (2014) The anatomy and physics of ZDR columns: investigating a polarimetric radar signature with a spectral bin microphysical model. *Journal of Applied Meteorology and Climatology*, 53, 1820–1843.

Lakshmanan, V., Hondl, K., Potvin, C.K. and Preignitz, D. (2013) An improved method for estimating radar echo-top height. *Weather and Forecasting*, 28, 481–488.

Liu, Z., Marchand, R. and Ackerman, T. (2010) A comparison of observations in the tropical western Pacific from ground-based and satellite millimeter-wavelength cloud radars. *Journal of Geophysical Research; Atmospheres*, 115(D24). https://doi.org/10.1029/2009JD013575

Lock, A.P., Brown, A.R., Bush, M.R., Martin, G.M. and Smith, R.N.B. (2000) A new boundary layer mixing scheme. Part I: scheme description and single-column model tests. *Monthly Weather Review*, 128, 3187–3199.

Mason, P.J. (1994) Large-eddy simulation: a critical review of the technique. *Quarterly Journal of the Royal Meteorological Society*, 120, 1–26.

Miltenberger, A.K., Field, P.R., Hill, A.A., Rosenberg, P., Shipway, B.J., Wilkinson, J.M., Scovell, R.W. and Blyth, A.M. (2018) Aerosol–cloud interactions in mixed-phase convective clouds – Part 1: aerosol perturbations. *Atmospheric Chemistry and Physics*, 18, 3119–3145. https://doi.org/10.5194/acp-18-3119-2018

Nicol, J., Hogan, R.J., Stein, T.H.M., Hanley, K.E., Clark, P.A., Halliwell, C.E., Lean, H.W. and Plant, R.S. (2015) Convective updraught evaluation in high-resolution NWP simulations using single-Doppler radar measurements. *Quarterly Journal of the Royal Meteorological Society*, 141, 3177–3189.

Pearson, K.J., Hogan, R.J., Allan, R.P., Lister, G.M.S. and Holloway, C.E. (2010) Evaluation of the model representation of the evolution of convective systems using satellite observations of outgoing longwave radiation. *Journal of Geophysical Research; Atmospheres*, 115(D20). https://doi.org/10.1029/2010JD014265

Plummer, D.M., French, J.R., Leon, D.C., Blyth, A.M., Lasher-Trapp, S., Bennett, L.J., Dufton, D.R.L., Jackson, R.C. and Neely, R.R. (2018) Radar-derived structural and precipitation characteristics of ZDR columns within warm-season convection over the United Kingdom. *Journal of Applied Meteorology and Climatology*, 57, 2485–2505.

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

Scovell, R.W. and Al-Sakka, H. (2016) A point cloud method for retrieval of high-resolution 3D gridded reflectivity from weather

radar networks for air traffic management. *Journal of Atmospheric and Oceanic Technology*, 33, 461–479.

Skamarock, W.C. (2004) Evaluating mesoscale NWP models using kinetic energy spectra. *Monthly Weather Review*, 132, 3019–3032.

Stein, T.H.M., Hogan, R.J., Hanley, K.E., Nicol, J.C., Lean, H.W., Plant, R.S., Clark, P.A. and Halliwell, C.E. (2014) The three-dimensional morphology of simulated and observed convective storms over southern England. *Monthly Weather Review*, 142, 3264–3283. https://doi.org/10.1175/MWR-D-13-00372.1

Stein, T.H.M., Hogan, R.J., Clark, P.A., Halliwell, C.E., Hanley, K.E., Lean, H.W., Nicol, J.C. and Plant, R.S. (2015) The DYMECS project: a statistical approach for the evaluation of convective storms in high-resolution NWP models. *Bulletin of the American Meteorological Society*, 96, 939–951.

Stein, T.H.M., Keat, W., Maidment, R.I., Landman, S., Becker, E., Boyd, D.F.A., Bodas-Salcedo, A., Pankiewicz, G. and Webster, S. (2019) An evaluation of clouds and precipitation in convection-permitting forecasts for South Africa. *Weather and Forecasting*, 34, 233–254.

Testud, J., Le Bouar, E., Obligis, E. and Ali-Mehenni, M. (2000) The rain profiling algorithm applied to polarimetric weather radar. *Journal of Atmospheric and Oceanic Technology*, 17, 332–356. https://doi.org/10.1175/1520-0426(2000)017<0332:TRPAAT>2.0.CO;2

Warren, R.A. and Protat, A. (2019) Should interpolation of radar reflectivity be performed in Z or dBZ?. *Journal of Atmospheric and Oceanic Technology*, 36, 1143–1156.

Zhang, J., Howard, K. and Gourley, J. (2005) Constructing three-dimensional multiple-radar reflectivity mosaics: examples of convective storms and stratiform rain echoes. *Journal of Atmospheric and Oceanic Technology*, 22, 30–42.