METEOROLOGICAL APPLICATIONS
Meteorol. Appl. **16**: 425–432 (2009)
Published online 28 May 2009 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/met.150

RMetS
Royal Meteorological Society

A method for using weather radar data to test cloud resolving models

Peter T. May[a]* and Todd P. Lane[b]

[a] *Centre for Australian Weather and Climate Research (CAWCR) - A partnership between the Australian Bureau of Meteorology and CSIRO, GPO Box 1289, Melbourne, 3001, Victoria, Australia*

[b] *School of Earth Sciences, The University of Melbourne, Parkville, Victoria, Australia*

ABSTRACT: A simple method for extracting useful metrics to characterize cloud fields from volumetric radar data is described along with the use of these metrics to evaluate the performance of cloud resolving models. This method involves the calculation of a 'Statistical Coverage Product', where the fraction of a grid of radar data covered by various reflectivity thresholds and microphysical classifications is calculated, along with profiles of maximum reflectivity, areas of convective and stratiform rain, and the amounts of rainfall in these categories. The application of this method is illustrated, whereby radar observation and a cloud resolving model simulation are compared for a tropical island thunderstorm near Darwin, Australia. Copyright © 2009 Royal Meteorological Society

KEY WORDS weather radar for clouds; model validation with radar; model verification using radar

Received 16 November 2008; Accepted 25 March 2009

1. Introduction

Cloud resolving models (CRM) and high resolution mesoscale models are widely used research and operational tools. They are used to provide dynamically complete and consistent data for understanding complex meteorological systems, parameterization development and understanding physical processes (e.g. Tao and Simpson, 1989, and many others). They have also been included within climate models as a substitute for parameterizing convective and cloud processes (Randall *et al*., 2003). Furthermore, operational mesoscale models are approaching true cloud resolving resolution. These wide applications underpin the need to develop quantitative methodologies for assessing the success of the models in simulating real weather systems, quantifying model performance and diagnosing limitations.

There has been a number of approaches described in the recent literature that aim to compare cloud observations to model simulations. For example, as part of the CLOUDNET project, Illingworth *et al*. (2007) use a statistical approach, which compares profiles from cloud observatories and numerical models to examine cloud cover as a function of time and height after suitable averaging over time has been undertaken. A more sophisticated approach, to examine data from model columns and profile observations, is to compare probability distributions. For example, Jakob *et al*. (2004) randomly sampled model columns and compared these with observations on a statistical basis. Other approaches include the comparison of joint height/reflectivity distributions (so called CFADs, Yuter and Houze, 1995) of radar reflectivity and vertical motion associated with models and observations (Rogers *et al*., 2007). However, approaches such as those outlined by Jakob *et al*. lose information associated with temporal structure.

Sensitive scanning radar offers the additional opportunity to sample 3D fields of condensate with high time resolution, at a similar spatial resolution to CRM fields (e.g. Zeng *et al*., 2001) reducing differences due to representativeness errors (Tustison *et al*., 2001). There are many potential quantitative approaches that can take advantage of the 4D data. For example, Tuttle *et al*. (1989) examined maximum observed reflectivity as a function of height and time, and compared the reflectivity to simulated cloud properties. However, to date most radar data comparisons with CRMs have typically been qualitative, where the CRM fields are assessed based on whether they have a similar appearance to the observations. Furthermore, many such comparisons have been limited because they do not compare similar measures of cloud coverage.

Here, a simple method for quantitatively comparing simulations of cloud fields of precipitating systems with radar data is described: the method is robust against measurement error and preserves time and height information. This method involves the calculation of a 'Statistical Coverage Product' (SCP) where the fraction of a grid of radar and model data covered by various reflectivity thresholds and microphysical classifications is calculated along with other parameters such as profiles of maximum

* Correspondence to: Peter T. May, Centre for Australian Weather and Climate Research (CAWCR) - A partnership between the Australian Bureau of Meteorology and CSIRO, GPO Box 1289, Melbourne, 3001, Victoria, Australia. E-mail: p.may@bom.gov.au

reflectivity and areas of convective and stratiform rain and amounts. The SCP represents an improvement over previous methods because it is designed to create products from models and observations that can be directly compared. These comparisons can discriminate between areas of low or high reflectivity, as well as convective or stratiform rain. Thus, the SCP has the potential to assess many aspects of CRM performance.

The approach for model validation will be illustrated using radar observations and modelling results for a typical afternoon thunderstorm over the Tiwi Islands north of Darwin, Australia. These are intense thunderstorms that form on an almost daily basis on these islands and have been extensive studies observationally (e.g. Keenan *et al*., 2000; Wilson *et al*., 2001) and with models (e.g. Golding, 1993; Crook, 2001). The model field used herein is an example from simulations aimed at the study of convectively generated gravity waves and is not tuned for optimal cloud structure (Lane and Reeder, 2001). This simulation is for illustrative purposes rather than delving in detail into potential model improvements, but the use of a less than optimized simulation for cloud structure is highly illustrative of the potential of these tools. The example will include qualitative diagnostics and quantitative measures.

The remainder of this note is organized as follows. Section 2 describes the two datasets that will be compared: the radar observations and the CRM simulation. Section 3 outlines the methodology for calculating the SCP, the results of which are presented in Section 4 along with a comparison of the radar and model data. Possible future applications of the SCP are detailed in Section 5, along with the conclusions.

2. Data for comparison

2.1. Radar observations

This study uses data from a 5.5 cm wavelength scanning polarimetric weather radar system located at Gunn Point (12.25 °S, 131.04 °E) near Darwin, Australia (C-Pol: Keenan *et al*., 1998), although much of the method can be applied to conventional weather radar. The location of C-Pol is depicted in Figure 1. Detailed descriptions of the enhanced capabilities associated with such polarimetric radar systems can be found in Zrnic and Ryzhkov (1999), but the germane issue here is their capability for much greater accuracy of rainfall estimates compared with conventional radar and the ability to describe the microphysical characteristics of the precipitation. The polarimetric radar measures several variables in addition to the reflectivity at horizontal polarization (Z_H). In particular, the difference in reflectivity at the two polarizations ($Z_{DR} = Z_H/Z_V$ where Z_V is the reflectivity at vertical polarization) provides information on the mean rain drop size, which is at the core of rainfall measurement uncertainties, and the differential phase on propagation (Φ_{DP}) that provides a measure of the attenuation of the signals. The gradient of the Φ_{DP}, the specific differential phase, K_{DP}, also provides an attenuation free estimate of the rainfall rate, which is particularly useful in high rain rate situations or when mixed phase precipitation is present. The final parameter measured by the C-Pol radar is the cross-correlation of the signal at the two polarizations ($\rho_{HV}(0)$), which supplies evidence of mixtures of precipitation type. These capabilities have been discussed at length in many papers (e.g. May *et al*., 1999; Zrnic and Ryzkhov, 1999; Bringi *et al*., 2002).

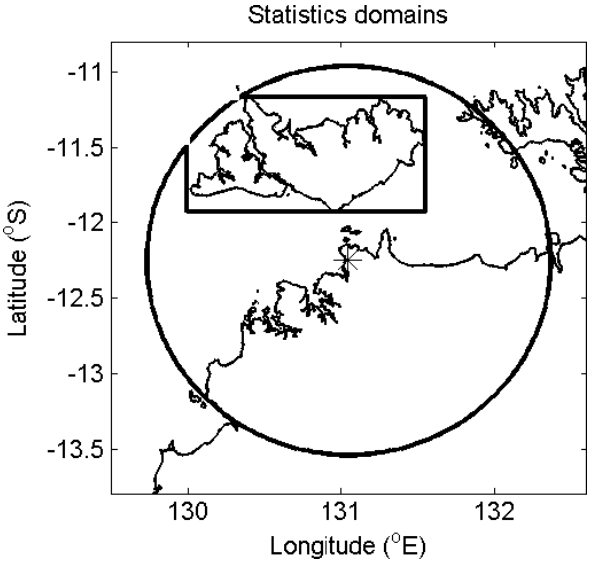

Figure 1. Location of the the C-Pol radar site at Gunn Point (marked by a star), and the rectangular domain used to calculate statistics. The circle marks the 150 km range of the radar data.

These additional observations lead to at least two significant capabilities compared with conventional radar. Firstly, it is possible to measure rainfall quantitatively more accurately. This is because one can use rainfall estimators optimized for different rainfall rates, including using information of the underlying dropsize distributions and correct for the effects of radar attenuation. The second major capability is the determination of hydrometeor species. Different hydrometeors occupy different parts of the [Z_H, Z_{DR}, K_{DP}, $\rho_{HV}(0)$, temperature] phase space so the combination of parameters can be used with good success (e.g. Keenan, 2003; May and Keenan, 2005). In addition to this, scatter from biological scatterers can be identified and separated from weather echoes (e.g. Schuur *et al*., 2003).

The radar observations that are used in this study were recorded on 10 February 2006. Example reflectivity cross-sections are shown in Figure 2: these will be described in more detail later in the paper.

2.2. CRM simulation

In order to illustrate the comparison methodology, CRM output from an idealized simulation of deep convective development in the region of interest is examined. Note that this simulation was not optimized to produce the best possible agreement with observations of the cloud

DOI: 10.1002/met

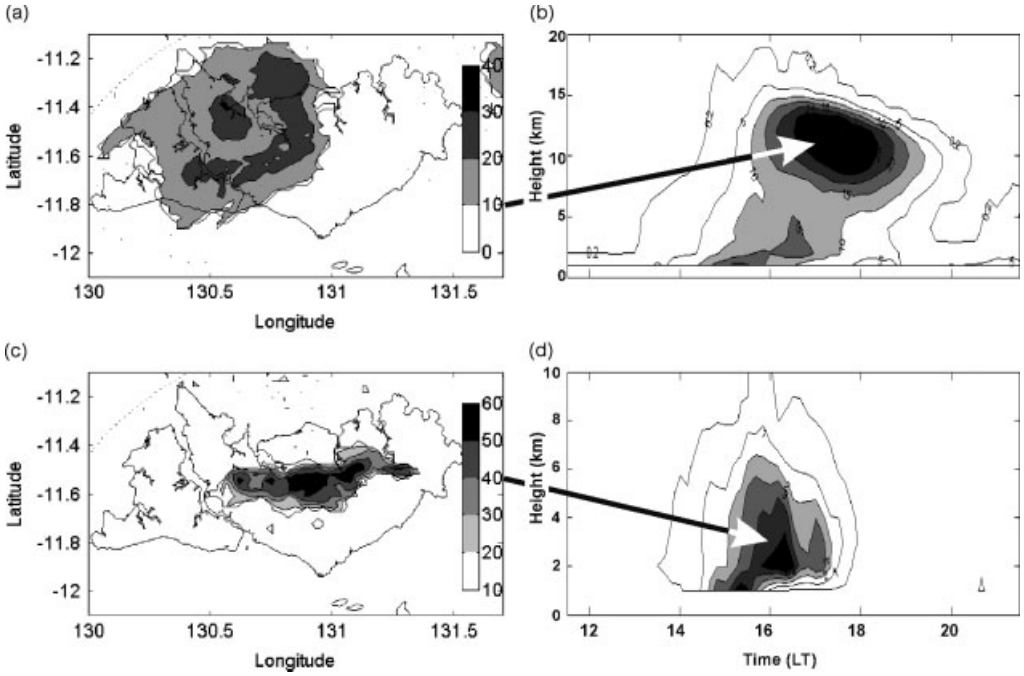

Figure 2. Examples of observed reflectivity fields and how they map onto the SCP product. The reflectivity fields are CAPPI (reflectivity fields at a constant altitude) for (a) 11 km (1720 LT) and (b) 3 km (1600 LT) on 10 February 2006. The 10 dBZ (c) and 40 dBZ (d) fraction at these times and heights are 34.7 and 4.4% respectively.

structures *per se*. Thus it will be an excellent example to highlight how the radar diagnostics can be used to highlight model deficiencies and suggest approaches for improvement.

The model simulation is a variation on that reported by Lane and Reeder (2001). It uses the Clark cloud model (Clark, 1979) over a model domain that is 300 × 300 km in the horizontal and 30 km high. The horizontal grid spacing is 1.5 km, and the vertical grid spacing varies from 25 m near the surface to 250 m further aloft. Explicit parameterizations of cloud microphysics is incorporated *via* a combination of the Kessler (1969) warm rain and the Koenig and Murray (1976) ice microphysics schemes. In this simulation, the topography of the Tiwi Islands is replaced by an idealized flat elliptical island that has a major axis in the east–west direction of 140 km and a minor axis in the north–south direction of 70 km. The model is initialized with a sounding released over the Tiwi Islands at 1330 UTC on 26 November 1995 and convection is forced by time-variant surface fluxes. From the simulated model fields, a forward model is used to derive the contribution of each microphysical class to the equivalent radar reflectivity assuming Rayleigh scatter.

The Kessler (1969) warm rain scheme assumes an exponential droplet distribution, such that its probability distribution follows:

$$N(D) = N_0 \mathrm{e}^{-\lambda D} \tag{1}$$

where D is the droplet diameter, N is the number of drops *per* unit volume in diameter range δD, and N_0 is a constant (equal to 10^7 m^{-4} in these simulations).

By integrating this distribution across all diameters, it can be shown that the total mass of rain per unit volume is:

$$M_\mathrm{R} = \pi \rho_\mathrm{l} N_0 / \lambda^4 \tag{2}$$

where ρ_l is the density of liquid water. The modelled rain water mixing ratio is:

$$q_\mathrm{R} = M_\mathrm{R} / \rho \tag{3}$$

where ρ is the density of air and it follows that $\lambda = (\pi \rho_\mathrm{l} N_0 / \rho q_\mathrm{R})^{1/4}$.

Kessler showed that the reflectivity factor for rain, Z_R, is equal to $Z_\mathrm{R} = 720 \times 10^{18} N_0 \lambda^{-7}$ mm^6 m^{-3} (when λ is in units of m^{-1}): Z_R is calculated from the simulation using the modelled values of q_R.

The Koenig and Murray (1976) scheme is a two-category double moment ice parameterization. The small ice category (Type A) is the result of heterogeneous nucleation and diffusional growth of ice particles. The large ice category (Type B) is the result of diffusional growth and the freezing of raindrops after collision with a Type A ice particle. For complete details of the implementation see Bruintjes *et al*. (1994), and for the purposes of this study we refer to these categories as simply Type A and Type B ice. For comparison with the radar, the Type A ice represents cloud ice particles while Type B captures snow and graupel that dominate the radar reflectivity in the regime that weather radars are sensitive. Each ice particle category is represented by the number concentration, N_I, and the mixing ratio, q_I, from which the ice mass *per* unit volume, M_I, can be calculated. For simplicity it is assumed that Type A ice does not

DOI: 10.1002/met

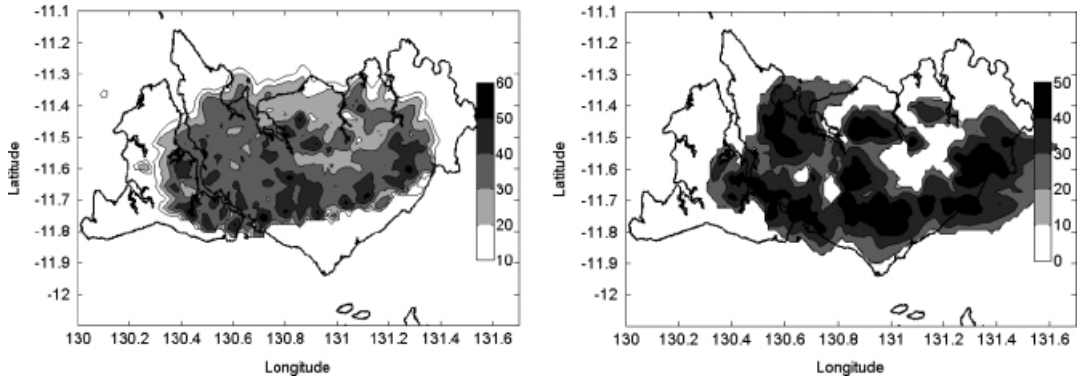

Figure 3. Equivalent reflectivity calculated from the simulated model fields at the time of storm system development approximately corresponding to the radar images in Figure 2. Model reflectively at 2.5 km (a) and 11 km (b) is shown, both at 1230 LT.

contribute to the reflectivity field since the particles are small and the reflectivity scales as the particle diameter to the sixth power, and that the Type B ice is completely dry. Following Ferrier (1994), who assumed an exponential distribution, the reflectivity contribution from dry ice is:

$$Z_I = 1.634 \times 10^4 (M_I^2 / N_I) \quad (4)$$

where Z_I, M_I, and N_I are in units of mm^6 m^{-3}, g m^{-3}, and #/litre, respectively.

Using the above relations, the total reflectivity from the CRM simulation in dBZ is 10 $\log_{10}$ $(Z_R + Z_I)$, which is analysed over the same domain of interest as the radar data. Examples of model-derived reflectivity cross-sections are shown in Figure 3: these are described in more detail later in the paper.

Of course, because of the approximations used in the CRM microphysics parameterizations, there is not a one-to-one correspondence between model microphysics and the radar derived classifications, yet some inferences can be made. First, the model fields can be used to calculate a radar reflectivity that allows direct comparison. In other instances there may be reasonable correspondences, although not exact. Here, the Type B ice is similar to the radar derived 'snow categorization' and it will be shown later that there is a close match between the model and observed rain fractions. There are also model fields that can be compared with observed proxies. For example the structure of the graupel and hail areas estimated by radar can be compared with model convective updraft area (here defined as $w > 5$ m s^{-1} based on observations of hail production with updrafts of this magnitude in tropical showers by May *et al*. (2002) and May and Keenan (2005). Observational studies have found that updrafts of this magnitude near the freezing level generally correspond to increased probability of hail production and lightning (e.g. May and Keenan, 2005).

3. A statistical coverage product (SCP)

Radar data sets are large and for many applications are cumbersome to use in raw form. However, there is a number of approaches available to generate derived products from the data that are of immediate use both for meteorological analysis and model evaluation. The approach described herein is in some way similar to the CLOUDNET approach (Illingworth *et al*., 2007), but extended to three dimensional data. The most straightforward extension of the cloud profile analysis might be to construct a three-dimensional cloud mask based on gridded radar reflectivity data. Note that such a cloud mask using weather radar data is limited by the radar sensitivity, which for the C-POL radar is ~0 dBZ at the maximum range of 150 km, compared with cloud radars which often measure echoes significantly weaker than −30 dBZ. However, such a dataset would not be substantially smaller than gridded reflectivity and classification data, and the relatively poor resolution at long ranges and gaps in the vertical sampling of the volume scanning would cause additional difficulties. Nevertheless, these ideas can be built on to obtain height resolved statistics from gridded data. A simple approach is to calculate the fraction of a domain that is covered by radar echoes greater than some nominal reflectivity threshold for a range of thresholds. Also, the fractional coverage of each microphysical classification, and the height profile of the maximum reflectivity within the domain, can also be determined. Thus the method is as follows:

1. interpolate the radar reflectivity and microphysical classifications from the polarimetric radar data onto a Cartesian grid and run a forward reflectivity calculation on the model fields to produce a matching grid, and,
2. at each altitude level of the grid calculate:
   (a) the fraction of the total area covered by Z > [10, 20, 30, 40] dBZ;
   (b) the fraction of the total area covered by rain, snow, graupel, hail, and,
   (c) the maximum reflectivity anywhere in the grid.

Note that 2(a) and (b) effectively represent probabilities and that (c) is a proxy for the maximum intensity of the cloud system that has been used before for model validation (Tuttle *et al*., 1989). Also, note that 2(a) and

DOI: 10.1002/met

2(c) can be applied to any scanning radar system. Furthermore typical calibration and measurement uncertainties of ~1 dBZ will not substantially affect these statistics.

These new diagnostics can be simply augmented with other parameters that preserve the temporal information and taking advantage of the similar resolutions. For example, it is convenient to calculate the area total rainfall as well as the convective and stratiform contributions to the average rain and the relative areas and mean rain rates of the convective and stratiform rain from both the radar and model fields. Here, the convective and stratiform areas are separated using the texture based algorithm described by Steiner *et al.* (1995). The rain rates are converted to a pseudo reflectivity, Z, using standard Z-R relations for this. In this algorithm, a pixel is classed as convective if one of the following applies.

1. $Z > 40$ dBZ, or,
2. Calculate the average reflectivity in a 10 km diameter circle about the pixel, Z_{BG}, and if $\Delta Z = Z - Z_{BG}$ satisfies:

$$\Delta Z \geq \begin{cases} 10 & Z_{BG} < 0 \\ 10 - {Z_{BG}}^2/180 & 0 \leq Z_{BG} < 42.43 \\ 0 & Z_{BG} > 42.43 \end{cases} \quad (5)$$

If 1 or 2 is satisfied, then the surrounding pixels to a radius R are also defined as convective where $R = 1$ km for $Z < 25$ dBZ, and then increases linearly with reflectivity until $R = 5$ km around a 40 dBZ pixel.

These approaches can be applied over the whole radar and model domain or in particular areas for specific applications. Weather radar volume coverage is approximately a current GCM grid box size so that radar statistics gathered on this scale have direct relevance to larger scale modelling, for example for comparisons with parameterized convection (convective fraction). The approach can also be applied to more limited spatial domains to match models of higher resolution or to focus on particular systems.

In the next section, the method will be illustrated with comparisons between modelled and observed storms over the Tiwi Islands north of Darwin, Australia. The domain used for the comparison is illustrated in Figure 1. The purpose of this comparison is to illustrate the features of the technique, using a simulation with some significant shortcomings, rather than to criticize a particular model.

4. Results and application

Before applying the SCP analysis, first consider qualitative comparisons between 'snap-shots' of the observed and simulated reflectivity fields (Figures 2 and 3). The times are chosen so that the respective storms are in a mature stage of the system evolution. This difference in the model and observed times already highlight one weakness of the model: the tendency for convection to form far too early in the day. This may be because the 1.5 km resolution of the model prohibits the production of small-scale boundary layer cumulus, resulting in an unphysical rate of convective development at low-levels. Apart from this limitation, the simulation is qualitatively reasonable with areas of deep, intense convective cloud similar to those observed, although the model reflectivities above the freezing level are excessive. However, there are some obvious differences. These differences include the wider area of reflectivity between 20 and 30 dBZ at low altitudes in the model simulation that is not seen in the observations, and the high, relatively peaked distribution of reflectivity at the altitude of the anvils. However, convective systems are often highly stochastic in their detailed structure so that detailed intercomparison of the modelled and observed field at a specific height and time is of limited use.

The SCP fields for both the model and a typical tropical island thunderstorm case are shown in Figure 4. As mentioned in Section 2, the modelled Type B ice is compared to the radar snow category, and the modelled updraft area (>5 m s^{-1}) is compared to the hail area. The features of the SCP fields elucidate many of the consistencies and inconsistencies between the model and the radar observations. It is apparent from the 40 and 10 dBZ areas that in both the model and the simulations the maximum in deep convection activity is followed by anvil production. However, these fields highlight the excessive reflectivities in the simulated ice phase, with even the anvil producing echoes in excess of 40 dBZ. Whereas, in the observations, the anvil is really only visible in the 10 (and 20, not shown) dBZ SCP. Moreover, the echo top of the anvil in the model decreases in altitude more slowly than observed, suggesting that the ice problem may be associated with very high concentrations rather than excessive size. This is supported by the much better agreement in the snow/ice particle areas.

The SCP fields in Figure 4 can also be quantitatively compared. To illustrate this the method described by Roberts and Lean (2008) that has the potential to objectively assess model performance with regard to forecast probabilities is adapted. The approach is to calculate a Fractions Skill Score (FSS), which is a variation on the Brier skill score. Following Roberts and Lean, the FSS is given by:

$$FSS = 1 - \frac{FBS}{\frac{1}{N}[\sum_{j=1}^{N} p_j^2 + \sum_{j=1}^{N} o_j^2]} \quad (6)$$

where $0 \leq p_j \leq 1$ is the forecast fraction, $0 \leq o_j \leq 1$ is the observed fraction, and,

$$FBS = \tfrac{1}{N} \sum_{j=1}^{N} (p_j - o_j)^2 \quad (7)$$

DOI: 10.1002/met

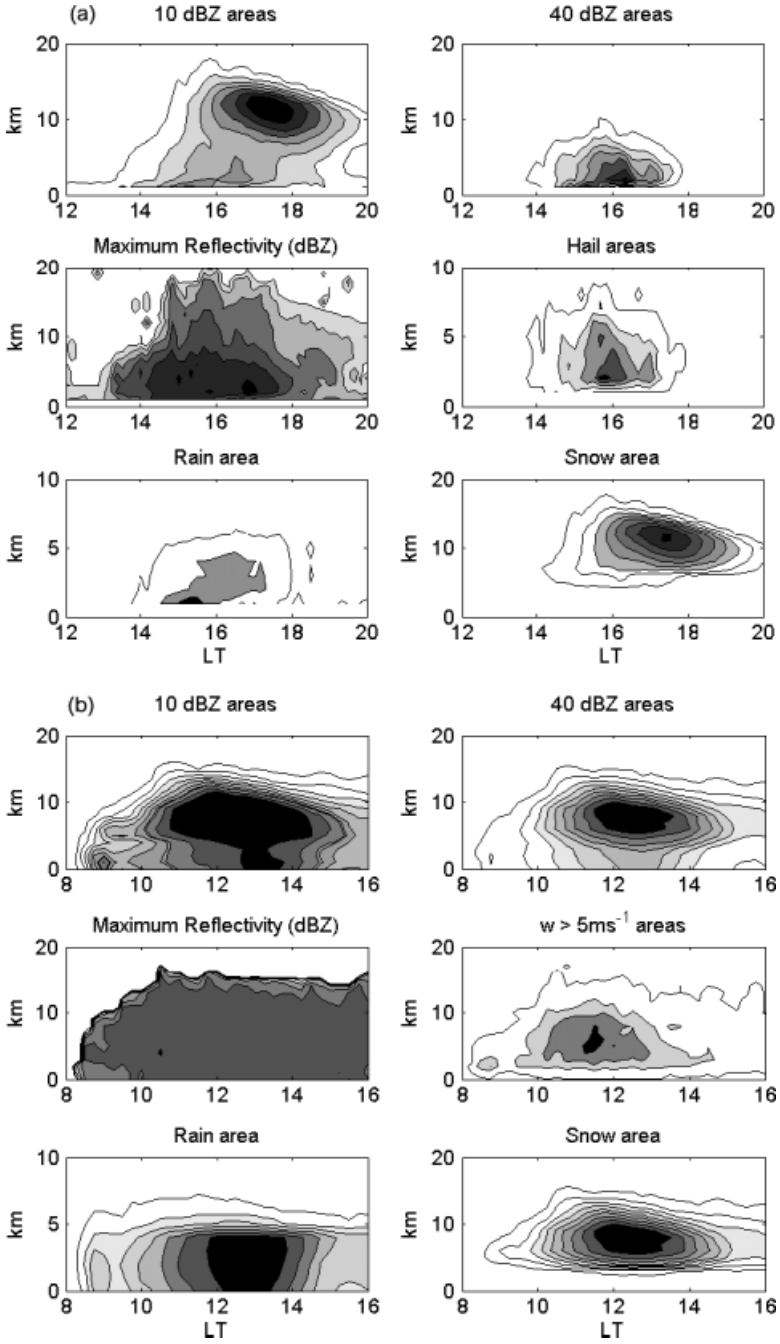

Figure 4. Time/height series of the SC products for the Tiwi Islands domain from C-Pol radar observations from 10 February 2006 (a) and corresponding statistics from the cloud resolving model (b). Times are local. Contour intervals: 10, 40 dBZ, rain, snow is 5% outline at 1%. Maximum reflectivity is contoured every 10 dBZ from 0 and hail and updraft areas are contoured every 0.5%.

is a version of the Brier score where fractions are compared.

The normalization of the FSS is by the worst possible value of FBS in the data where there is no co-location. The FSS has a range of [0, 1] where 0 is a complete mismatch and 1 is perfect. It is sensitive to rare events and gives a value of 0 if no events are forecast or observed.

The errors associated with timing ensure that the FSS $\sim$0, but in order to examine the ability of the model to simulate realistic cloud property patterns, a time shift of 4 h has been inserted into the model analysis to bring the period of maximum convection into phase with observations. Some sample values of the FSS are given in Table I where the radar data has been sampled once per hour to match the model output. The FSS values are fairly insensitive to the exact start time of the radar estimate with standard deviations of the FSS estimates of only about 0.01 for radar start points $+/-30$ min, except for the snow area which has about twice the sensitivity.

This example is only meant for illustrative purposes, but clearly exposes the greater difficulty of predicting the details of the convective cores with respect to the general precipitation field. While the areal coverage of radar echoes is reasonably well represented with a FSS

DOI: 10.1002/met

Table I. Fractions Skill Score (FSS) for a number of the model fields compared with observations. The model time has been shifted for optimal agreement. The statistics were calculated for all heights except where noted and the radar data re-sampled to the model time resolution.

Parameter	FSS value
10 dBZ fraction	0.53
40 dBZ fraction	0.03
Rain (up to 8 km) fraction	0.20
Snow/Type B ice fraction	0.57

of ∼0.5, the high reflectivity values are very low. This is not only because of the gross over-estimation of radar reflectivity in the anvil evident in Figure 4. If the FSS is calculated using data only up to an altitude of 5 km, the resulting FSS is still only 0.07. The actual anvil area is reasonably well represented with high values of the snow FSS.

From these comparisons it is clear that there is a gross overestimation of large ice particle (reflectivity) by the model. This could simply be due to inadequacies in the forward model, but is most likely due to the relatively simple nature of the treatment of ice formation. In this two-category scheme the ice must to be either large or small, i.e. there are no intermediate sizes. This should lead to overly large ice (Type B) particle concentrations of larger particles that make a substantial contribution to the reflectivity field. It is expected that more sophisticated ice schemes that allow for intermediate ice sizes would perform better than the scheme presented here.

As demonstrated above, the SCP offers a useful way of augmenting simple qualitative comparisons, and provides quantitative measures as well as qualitative insights into more specific model limitations. The SCP results can also be aggregated over many cases and domains to investigate systematic errors. Therefore, the SCP product has clear and straightforward applications for the assessment of cloud resolving, mesoscale and single column modelling.

As discussed above, there are many other complementary approaches for the validation of the model data. For example, area averages of near surface observations of rain may also be used for testing the model. Here the convective and stratiform rain amounts and areas have been disaggregated using the texture based algorithm of Steiner *et al*. (1995) (as discussed in Section 3). Figure 5 shows time series of the observed and modelled area averaged total rainfall and the convective rain components of this average. The stratiform component is the difference. This clearly shows the super-abundance of the amount of stratiform rain in the model simulation. This analysis can then be augmented by the area of rainfall, so that both the area and intensity information is captured. Here the excessive area of stratiform rain is clearly observed. This problem is illustrated in Figure 3, with the area of lower reflectivities filling in between the convective cores and extending over much of the islands. The simulations

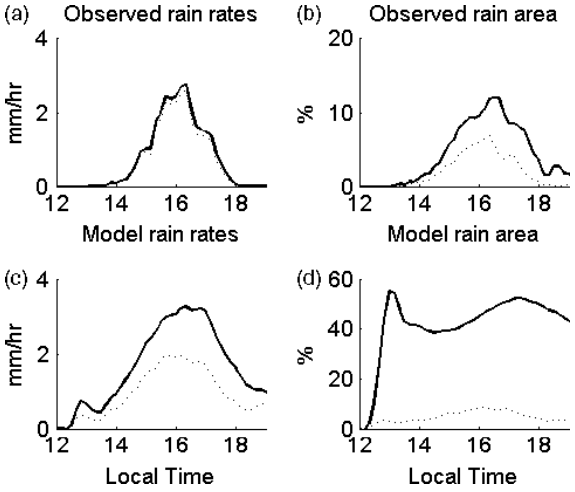

Figure 5. Time series of the domain average total rain rate (solid) and convective rain rate (dotted) for the radar observations (a, b) and model simulations (c, d). The time axis has been shifted by 4 h for the models to bring them into reasonable phase agreement as discussed in the text.

have this erroneous feature early in the simulation and maintained it throughout. In addition to this simple comparison, these derived statistics of rain rates and areas can also be compared both qualitatively and using standard metrics such as RMS errors.

Finally, for longer data records that include many events, approaches using probability scores and diagrams can be employed (e.g. Jakob *et al*., 2004). A summary of such verification methodologies can be found on-line at: http://www.bom.gov.au/bmrc/wefor/staff/eee/verif/verif_web_page.html

5. Conclusions

This note has described a simple procedure for reducing weather radar and model data sets for the purposes of analysis and validation that retains temporal structure and facilitates diagnostic and quantitative applications. The focus here has been on validation, but the metrics described can also be useful in the analysis of complex precipitation systems. A case study where the storms are strongly controlled by local circulations so that reasonable forecasts are expected has been shown. However, the method clearly highlighted model deficiencies and allowed diagnosis of areas where the improvements in the model physics can be found. The approach is equally valid in other situations and can also be applied to mesoscale models and in particular models carrying explicit microphysics as is becoming increasingly common. The described approach is even more appropriate for cases that are expected to be weakly forced and have a more stochastic character. The analysed fields can easily be extended to include variables such as rain water and precipitating ice content or the SCM parameters within just convective/stratiform areas. Radar analysis using these metrics is also likely to be useful in the formulation and testing of parameterizations in larger scale models.

DOI: 10.1002/met

Acknowledgements

This work has been partially supported by the United States DOE Atmospheric Radiation Measurement (ARM) program and by the Australian Research Council Discovery Projects funding scheme (DP0770381). We would like to gratefully acknowledge many fruitful discussions with Beth Ebert and Christian Jakob and the assistance of reviewers.

References

Bringi VN, Huang G-J, Chandrasekar V, Gorgucci E. 2002. A methodology for estimating the parameters of a gamma raindrop size distribution model from polarimetric radar data: application to a squall-line event from the TRMM/Brazil campaign. *Journal of Atmospheric and Oceanic Technology* **19**: 633–645.

Bruintjes RT, Clark TL, Hall WD. 1994. Interactions between topographic airflow and cloud/precipitation development during the passage of a winter storm in Arizona. *Journal of Atmospheric Science* **51**: 48–67.

Clark TL. 1979. Numerical simulations with a three-dimensional cloud model: lateral boundary condition experiments and multicellular severe storm simulations. *Journal of Atmospheric Science* **36**: 2191–2215.

Crook NA. 2001. Understanding hector: the dynamics of island thunderstorms. *Monthly Weather Review* **129**: 1550–1563.

Ferrier BS. 1994. A double-moment multiple-phase four-class bulk ice scheme. Part I: description. *Journal of Atmospheric Science* **51**: 249–280.

Golding BW. 1993. A numerical investigation of tropical island thunderstorms. *Monthly Weather Review* **121**: 1417–1433.

Illingworth AJ, Hogan RJ, O'Connor EJ, Bouniol D, Brooks ME, Delanoë J, Donovan DP, Eastment JD, Gaussiat N, Goddard JWF, Haefflin M, Klein Baltink H, Krasnov OA, Pelon J, Piriou J-M, Protat A, Russchenberg HWJ, Seifert A, Tompkins AM, Van Zadelhoff G-J, Vinit F, Willén U, Wilson DR, Wrench CL. 2007. CLOUDNET. *Bulletin of the American Meteorological Society* **88**: 883–898.

Jakob C, Pincus R, Hannay C, Xu K-M. 2004. Use of cloud radar observations for model evaluation: a probabilistic approach. *Journal of Geophysical Research* **109**: D03202, DOI: 10.1029/2003JD003473.

Keenan TD. 2003. Hydrometeor classification with a C-band polarimetric radar. *Australian Meteorological Magazine* **52**: 23–31.

Keenan TD, Glasson K, Cummings F, Bird TS, Keeler J, Lutz J. 1998. The BMRC/NCAR C-Band polarimetric (C-POL) radar system. *Journal of Atmospheric and Oceanic Technology* **15**: 871–886.

Keenan T, Rutledge S, Carbone R, Wilson J, Takahashi T, May P, Tapper N, Platt M, Hacker J, Sekelsky S, Moncrieff M, Saito K, Holland G, Crook A, Gage K. 2000. The maritime continent thunderstorm experiment (MCTEX): overview and some results. *Bulletin of the American Meteorological Society* **81**: 2433–2455.

Kessler E. 1969. *On the Distribution and continuity of Water Substance in Atmospheric Circulations*. Meteorological Monographs No.32, American Meteorological Society: Boston, MA; 84.

Koenig LR, Murray FW. 1976. Ice-bearing cumulus cloud evolution: numerical simulation and general comparison against observations. *Journal of Applied Meteorology* **15**: 747–762.

Lane TP, Reeder MJ. 2001. Modelling the generation of gravity waves by a maritime continent thunderstorm. *Quarterly Journal of the Royal Meteorological Society* **127**: 2705–2724.

May PT, Jameson AR, Keenan TD, Johnston PE, Lucas C. 2002. Combined wind profiler/polarimetric radar studies of the vertical motion and microphysical characteristics of tropical sea breeze thunderstorms. *Monthly Weather Review* **130**: 2228–2239.

May PT, Keenan TD. 2005. Evaluation of microphysical retrievals from polarimetric radar with wind profiler data. *Journal of Applied Meteorology* **44**: 827–838.

May PT, Keenan TD, Zrnic DS, Carey LD, Rutledge SA. 1999. Polarimetric measurements of rain at a 5 cm wavelength. *Journal of Applied Meteorology* **38**: 750–765.

Randall D, Khairoutdinov M, Arakawa A, Grabowski W. 2003. Breaking the cloud paramaterisation deadlock. *Bulletin of the American Meteorological Society* **84**: 1547–1564.

Roberts NM, Lean HW. 2008. Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Monthly Weather Review* **136**: 78–97.

Rogers RF, Black ML, Chen SS, Black RA. 2007. An evaluation of microphysics fields from mesoscale model simulations of tropical cyclones, Part 1: comparisons with observations. *Journal of Atmospheric Science* **64**: 1811–1834.

Schuur T, Ryzhkov A, Heinselman P, Zrnic D, Burgess D, Scharfenberg K. 2003. *Observations and Classification of Echoes with the Polarimetric WSR-88D Radar*, NSSL report, October, 2003, 46.

Steiner M, Houze RA Jr, Yuter SE. 1995. Climatological characterization of three-dimensional storm structure from operational radar and rain gauge data. *Journal of Applied Meteorology* **34**: 1978–2007.

Tao W-K, Simpson J. 1989. Modelling study of a tropical squall-type convective line. *Journal of Atmospheric Science* **46**: 177–202.

Tustison B, Harris D, Foufoula-Georgiou E. 2001. Scale issues in verification of precipitation forecasts. *Journal of Geophysical Research* **106**(D11): 11775–11784.

Tuttle JD, Bringi VN, Orville HD, Kopp FJ. 1989. The multiparameter radar study of a microburst – comparison with model results. *Journal of Atmospheric Science* **46**: 601–620.

Wilson JW, Carbone RE, Tuttle JD, Keenan TD. 2001. Tropical island convection in the absence of significant topography. Part II: nowcasting storm evolution. *Monthly Weather Review* **129**: 1637–1655.

Yuter SE, Houze RA Jr. 1995. Three-dimensional kinematic and microphysical evolution of Florida cumulonimbus. Part II: frequency distributions of vertical velocity, reflectivity, and differential reflectivity. *Monthly Weather Review* **123**: 1941–1963.

Zeng Z, Yuter SE, Houze RA Jr, Kingsmill DE. 2001. Microphysics of the rapid development of heavy convective precipitation. *Monthly Weather Review* **129**: 1882–1904.

Zrnic DS, Ryzhkov A. 1999. Polarimetry for weather surveillance radars. *Bulletin of the American Meteorological Society* **80**: 389–406.