METEOROLOGICAL APPLICATIONS
*Meteorol. Appl.* **21**: 444–458 (2014)
Published online 5 March 2014 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/met.1439

RMetS
Royal Meteorological Society

# The influence of erroneous background, beam-blocking and microphysical non-linearity on the application of a four-dimensional variational Doppler radar data assimilation system for quantitative precipitation forecasts

Shao-Fan Chang,[a] Juanzhen Sun,[b] Yu-Chieng Liou,[a]* Sheng-Lun Tai[a] and Ching-Yu Yang[a]
[a] *Department of Atmospheric Sciences, National Central University, Jhongli City, Taiwan*
[b] *Research Applications Laboratory, National Center for Atmospheric Research, Boulder, CO, USA*

**ABSTRACT:** A series of observation system simulation experiments (OSSEs) and a real case study are conducted to investigate the application of the Doppler radar data assimilation technique for numerical model quantitative precipitation forecasts (QPFs). A four-dimensional variational Doppler radar analysis system (VDRAS) is adopted for all experiments. The first set of OSSEs demonstrates that when the background field contains the imperfect information predicted from a mesoscale model, the incorrect convective-scale perturbations in the background can result in spurious scattered precipitation. However, a smoothing procedure can be used to remove the fine structures from the primitive model output in order to avoid this over-prediction. Results from the second set of OSSEs indicate that the lack of low-elevation data owing to radar scan and/or beam blockage could significantly alter the retrieved low-level thermal and dynamical structures when a different number of data assimilation cycles is applied. These impacts could lower the rainfall forecast capability of the model. The third set of OSSEs shows that, when the rainwater is assimilated over a long assimilation window, the non-linearity embedded in the microphysical process could lead the minimization algorithm in a wrong direction, causing a further degradation of the rainfall prediction. However, using multiple short assimilation cycles produces better minimization and forecast results than those obtained with a single long cycle. A real case experiment based on data collected during Intensive Operation Period (IOP) #8 of the 2008 Southwest Monsoon Experiment (SoWMEX) is conducted to provide a verification of the conclusions obtained from OSSEs under a realistic framework.

KEY WORDS forecasting; remote sensing; modelling

*Received 31 March 2013; Revised 19 August 2013; Accepted 27 August 2013*

## 1. Introduction

In the past few decades, close attention has been paid to using the data assimilation technique to provide proper initial conditions needed for weather prediction models, and to further improve the model spin-up time and forecast accuracy. Doppler radar can provide observation data with high temporal and spatial resolution that is suitable for convective scale research. A variety of radar data assimilation methods have been developed to improve the performance of numerical model prediction for convective systems. Gal-Chen (1978) first tried to infer the unobserved thermodynamic parameters (i.e. pressure and temperature) from the radar-derived three-dimensional winds. Since then, many studies have been carried out to use this concept to optimize the initialization of numerical models based on products retrieved from radar observations (Lin *et al.*, 1993; Crook, 1994; Crook and Tuttle, 1994; Weygandt *et al.*, 2002; Zhao *et al.*, 2006).

The three-dimensional variational data assimilation (3DVar) technique has been employed widely to assimilate Doppler radar data in severe weather studies (e.g. Xiao *et al.*, 2005; Hu *et al.*, 2006a, 2006b; Xiao and Sun, 2007; Chung *et al.*, 2009; Gao and Stensrud, 2012). The four-dimensional variational data assimilation (4DVar) adjoint technique is a more sophisticated algorithm, which considers the model trajectory over the assimilation window. Examples of 4DVar-based radar assimilation systems include the variational Doppler radar analysis system (VDRAS, Sun and Crook, 1997), Japan Meteorological Agency's 4DVar (Takuya *et al.*, 2011), and Weather Research and Forecasting (WRF) 4DVar (Sun and Wang, 2013; Wang *et al.*, 2013).

The use of the ensemble Kalman filter (EnKF) technique to assimilate Doppler and/or dual-polarimetric radar observations has also been investigated in numerous studies (Snyder and Zhang, 2003; Tong and Xue, 2005; Xue *et al.*, 2006; Jung *et al.*, 2008a, 2008b). The distinguishing feature of EnKF is its capability to provide flow-dependent covariance using ensemble forecast error statistics. Recently, active efforts have been made to develop a hybrid technique combining EnKF and 3DVar for the assimilation of radar data (Li *et al.*, 2012; Pan *et al.*, 2012).

The main purpose of this work is to discuss three important topics regarding the application of Doppler radar data assimilation technique to quantitative precipitation forecasts (QPFs). They include the treatment of a background field containing model prediction error, the lack of low-elevation radar observations owing to standard plan position indicator (PPI) scans and/or beam-blockage and the impact from microphysical non-linearity. The discussion of these three issues provides us a guideline to design an optimal assimilation strategy when both radar radial wind and reflectivity are available at multiple time

* Correspondence: Y.-C. Liou, Department of Atmospheric Sciences, National Central University, 32001 Jhongli City, Taiwan.
E-mail: ncuycliou@gmail.com

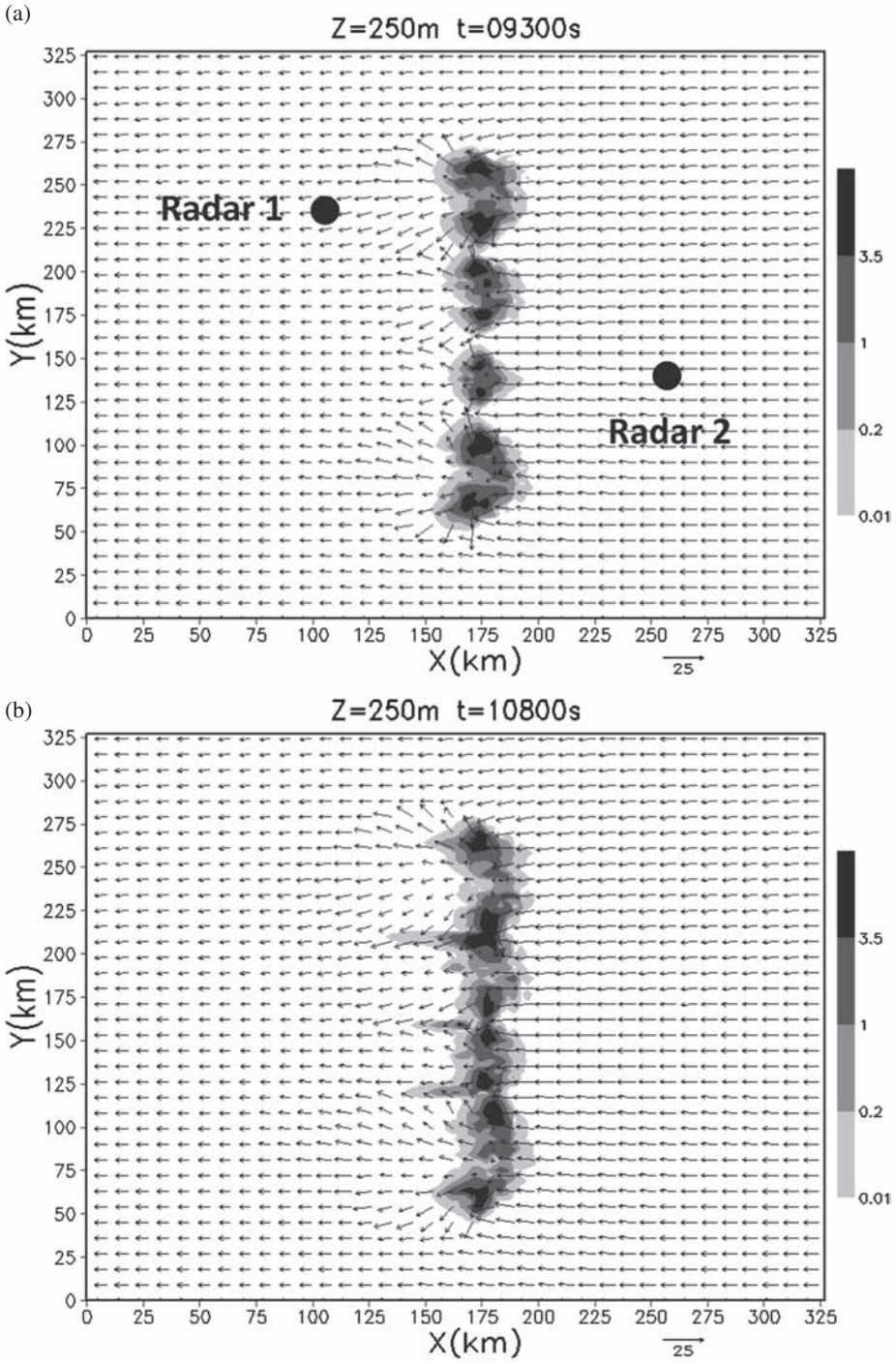


Figure 1. Horizontal cross section at the lowest level (250 m) showing the rainwater mixing ratio (shading) in g kg$^{-1}$ and flow vectors from the natural run: (a) at $T = 9300$ s and (b) 10 800 s. Vectors are plotted every 9 km. The positions of two pseudo-radars are marked by solid circles in (a).

levels. The radar data assimilation system adopted for all experiments is the VDRAS developed by the National Center for Atmospheric Research (NCAR). VDRAS uses the 4DVar technique to assimilate radial wind and reflectivity from single or multiple Doppler radar(s) to retrieve the unobserved three-dimensional winds, thermodynamic field and microphysics for the convective scale systems. VDRAS has been implemented for the assimilation of high-resolution data from operational Doppler radar and surface networks to produce real-time low-level analyses (Sun and Crook, 2001). VDRAS has also been used to demonstrate its potential for the short-term forecasting of severe storms in field projects (Sun, 2005; Sun and Zhang, 2008), Sydney 2000 Forecast Demonstration Project (Crook and Sun, 2002) and Beijing 2008 Forecast Demonstration Project (Sun *et al.*, 2010). Tai *et al.* (2011) were the first to apply VDRAS in Taiwan and its vicinity. They demonstrated that the model's QPF ability can be improved significantly by merging the VDRAS analysis fields with the WRF model.

This paper is organized as follows. In Section 2, the VDRAS system is briefly introduced. Section 3 describes the natural

 

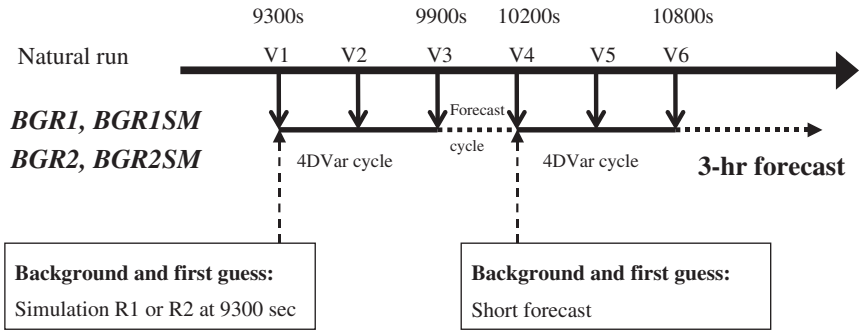


Figure 2. Illustration of VDRAS assimilation and forecast cycles for the experiments BGR1, BGR1SM, BGR2 and BGR2SM. The down-pointing arrows labelled V1-V6 at the top denote the assimilated volumes for both radar 1 and radar 2. The times at the top show the starting and finishing time for each 4DVAR cycle with respect to the simulation time for the natural run.

run and simulated radar observation for the observation system simulation experiments (OSSEs). The verification indices are defined in Section 4. In Section 5, the main topics are examined. A brief introduction of a real case study from Intensive Operation Period (IOP) #8 of the 2008 Southwest Monsoon Experiment (SoWMEX), the experimental design and the results are presented in Section 6. Finally, the summary and conclusions are presented in Section 7.

## 2. Description of VDRAS

VDRAS is composed of four major components: data acquisition, data pre-processing, 4DVar data assimilation and model outputs. Detailed descriptions can be found in Sun and Zhang (2008) and Sun *et al.* (2010).

The 4DVar radar assimilation system in VDRAS includes a forward cloud-resolving numerical model, the adjoint of the numerical model, a cost function and a minimization algorithm. The numerical model is formulated within a Cartesian coordinate system with a flat surface, and uses Kessler-type warm-rain microphysical parameterization. There are six prognostic variables in the numerical model: wind velocity components $(u, v, w)$, rain water mixing ratio $(q_r)$, total liquid water mixing ratio $(q_t)$ and liquid water potential temperature $(\theta_l)$. The total liquid water mixing ratio is the sum of rainwater, cloud water and the vapour mixing ratio. The temperature $(T)$ and the cloud water mixing ratio $(q_c)$ are diagnosed from the prognostic variables by assuming that all vapour greater than the saturation value is converted to cloud water. The pressure $(p)$ is diagnosed through a Poisson equation. An open lateral boundary condition is used in the VDRAS numerical model. The along-beam and cross-beam components of the horizontal inflow are determined by a combination of the radar radial velocity observation and the background wind field, respectively. The horizontal outflow is extrapolated using the values at the closest two inner grid points. The top and bottom boundary conditions are set to zero for vertical velocity, and all other variables are defined such that their normal derivatives vanish. The detailed model set-up can also be found from Sun and Crook (1997, 2001).

The application of the 4DVar technique in VDRAS allows the use of prognostic equations as constraints to minimize the cost function. VDRAS assimilates the Doppler radar data and then finds an optimal initial condition through the minimization algorithm. The cost function $(J)$ for measuring the misfit between the model state variables and the radar observations can be written as:

$$J = (\mathbf{x}_o - \mathbf{x}_b)^{\mathbf{T}} \mathbf{B}^{-1} (\mathbf{x}_o - \mathbf{x}_b) + \sum_{\sigma,t} [\eta_v \left(V_r - V_r^o\right)^2 + \eta_q \left(q_r - q_r^o\right)^2] + J_p + J_{mb} \quad (1)$$

where $\mathbf{x}_o$ is the model state variable at the beginning of the current assimilation window, $\mathbf{x}_b$ is background field forecast from the previous cycle and $\mathbf{B}$ denotes the background covariance matrix. The second term in Equation 1 represents the discrepancy from the radar observations, and it is assumed that there is no spatial error correlation between observations. The summation is over space $(\sigma)$ and time $(t)$. The co-efficients $\eta_v$ and $\eta_q$ represent the inverse of the observation error variance for radial velocity and rainwater, respectively. The variables $V_r$ and $q_r$ are the model output of Doppler radial velocity and rainwater mixing ratio, and the variables $V_r^o$ and $q_r^o$ are their observational counterparts, respectively. $V_r$ is calculated from the model Cartesian velocity components using the following relation:

$$V_r = \frac{x - x_r}{r} u + \frac{y - y_r}{r} v + \frac{z - z_r}{r} (w - V_T) \quad (2)$$

where $r$ represents the distance between a grid point $(x,y,z)$ and the radar location $(x_r,y_r,z_r)$ and $V_T$ $(\mathrm{m\,s^{-1}})$ is the terminal velocity for rain. The terminal velocity is estimated using rainwater through the following relation (Sun and Crook, 1997):

$$V_T = 5.4(p_0/\overline{p})^{0.4} (\rho_a q_r)^{0.125} \quad (3)$$

where $\overline{p}$ is the base-state pressure, $p_0$ is the pressure at the ground and $\rho_a$ represents density of air. The rainwater mixing ratio $q_r$ is estimated from the radar reflectivity ($Z$ in dBZ) by using the formula (Sun and Crook, 1997):

$$Z = 43.1 + 17.5 \log_{10} (\rho_a q_r) \quad (4)$$

assuming a Marshall–Palmer raindrop size distribution. A spatial and temporal smoothness penalty term $(J_p)$ is applied in Equation 1 to force the minimization results to smoothly fit the observations. The final term $J_{mb}$ denotes a mesoscale background field. This term ensures that the 4DVar analysis does not drift too far from the larger-scale background in the radar data-void region. The mesoscale background field in VDRAS combines data from *in situ* observations such as radiosondes, profilers, surface networks, velocity-azimuthal-display (VAD) analyses, mesoscale model analyses and reanalysis data through an objective analysis. Sun and Zhang (2008) and Sun *et al.* (2010) discussed the combination procedure in details.

 

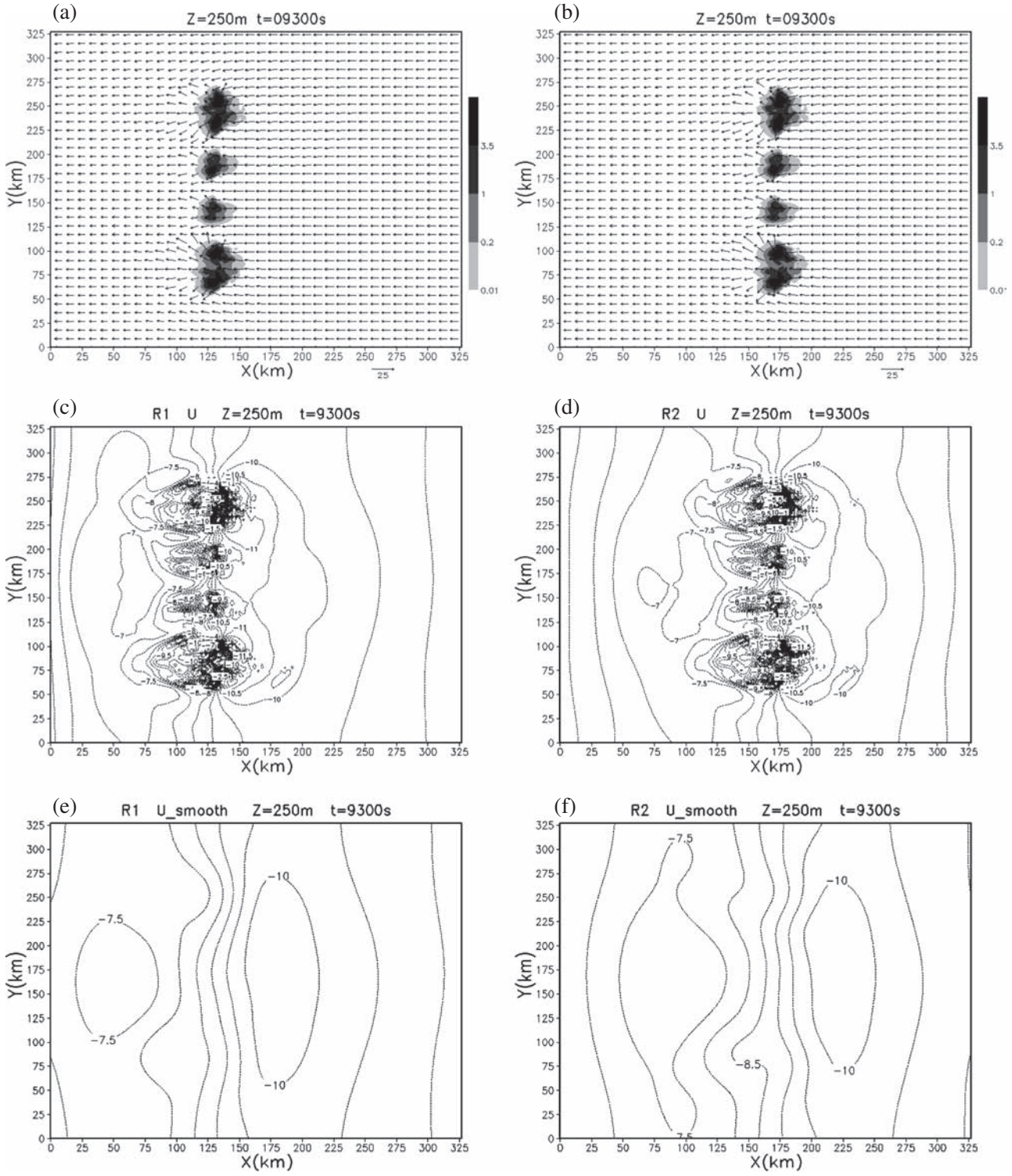


Figure 3. The horizontal cross section at $Z = 250$ m for (a, b) the rainwater mixing ratio in $g\,kg^{-1}$ and horizontal wind vectors shown every 9 km, (c, d) the $u$ component of wind (contour interval is $0.5\,m\,s^{-1}$) and (e, f) the $u$ component of wind smoothed by the Barnes technique (contour interval is $0.5\,m\,s^{-1}$) at a simulation time of $T = 9300$ s. The left and right columns are for simulations R1 and R2, respectively.

## 3. Natural run and simulated radar observation for the OSSEs

The natural run for the OSSEs is the simulation of a squall line using the cloud-resolving model built in VDRAS. The model domain used for both simulation and assimilation is $330 \times 330 \times 15\,km^3$. The grid spacing is 3 km in the horizontal and 500 m in the vertical, respectively. The time step for the model integration is 5 s.

The soundings used for the initial temperature and moisture profiles are from Weisman and Klemp (1982). This initial sounding can provide a suitable environment for convection development. The initial thermal perturbation is in a

 

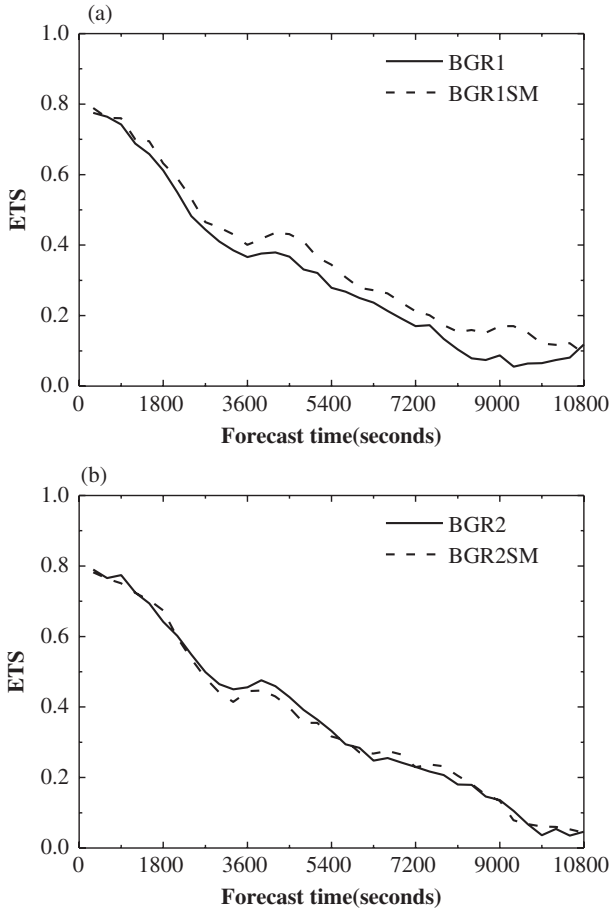


Figure 4. ETSs for predicted 5 min accumulated rainfall at threshold of 1.25 mm $(5\,\text{min})^{-1}$ *versus* forecast time for (a) BGR1 (solid line) and BGR1SM (dashed line), (b) BGR2 (solid line) and BGR2SM (dashed line).

north–south-oriented line with a 2.0 K maximum potential temperature excess (Klemp and Wilhelmson, 1978), superimposed by small (0.1 K) random perturbations. The unidirectional initial wind shear profile is surface-based (Weisman and Rotunno, 2004), with a magnitude of $15\,\text{m s}^{-1}$ within the lower 3 km. The initial $u$ component of wind at surface is $-10\,\text{m s}^{-1}$, while the $v$ component of wind at all levels is zero. The initial rainwater mixing ratio and vertical velocity are zero. Figure 1 parts (a) and (b) presents the horizontal distribution at $Z = 250$ m of the rainwater mixing ratio from the natural run, at $T = 9300$ and 10 800 s, respectively. Isolated convection cells appear during the early stage in the simulation (Figure 1(a)). In the mature stage, the individual cells grow and merge to form a squall line (Figure 1(b)).

The positions of two pseudo-radars are denoted in Figure 1(a). The maximum range of detection is set to 250 km. For the distance between radars, we adopt the average distance (166 km) among Taiwan's Central Weather Bureau (CWB) operational S-band Doppler radars. The two radars are set in the east and west of the squall line, respectively, with a distance of 170 km, to imitate an idealized scenario that the squall line can be completely observed under the coverage of two radars. The simulated observations of the radial velocity and reflectivity are calculated from the model outputs of the natural run through Equations 2 and 4, respectively. The reflectivity is available at each grid within the radar coverage. The radial velocity is considered observable only at the places where the reflectivity exceeds 0.0 dBZ.

## 4. Verification indices

The accuracy of the predicted accumulated rainfall is used to evaluate the model performance. In each time step ($\Delta T$), the accumulated precipitation $P$ (mm) for each grid at the lowest model plane can be written as:

$$P = \frac{\rho_a q_r V_T}{\rho_w} \times 1000 \times \Delta T \tag{5}$$

where $\rho_a$ and $\rho_w$ ($\text{kg m}^{-3}$) represent the density of air and water, respectively. On the basis of Equation 5, one can compute the required accumulated rainfall over a given period of time for verification.

For a quantitative comparison of the accuracy of the predicted precipitation obtained in different experiments, the equitable threat score (ETS) is chosen to verify the forecast performance. This index was proposed by Schaefer (1990) and Rogers *et al.* (1996), and can be defined as follows:

$$\text{ETS} = \frac{H - R}{F + O - H - R} \tag{6}$$

where $H$ stands for the number of correctly predicted points, $F$ is the number of model forecast points, $O$ denotes the number of observed points and $R$ represents the number of hits by chance, which is written as:

$$R = \frac{F \times O}{N} \tag{7}$$

where $N$ is the total number of points in the verification domain. The number of points is counted in $H$, $F$ and $O$ only when the precipitation is above a certain prescribed threshold. In this section, the threshold is 1.25 mm $(5\,\text{min})^{-1}$, which is a criterion converted from the official definition of a 'heavy rainfall event' by Taiwan CWB. When ETS reaches one, the forecast is considered perfect.

The root-mean-square error (RMSE) is also applied to conduct a quantitative comparison between the analysis from VDRAS and the observations, defined as follows:

$$\text{RMSE} = \sqrt{\frac{\sum (X_R - X_O)^2}{N}} \tag{8}$$

where subscripts $R$ and $O$ respectively represent the retrieved and observed values for a certain parameter $X$, and $N$ is the total number of grid points used for the computation. It should be noted that the observed data indicated in Equations 6–8 are the model outputs from natural run in all OSSEs.

## 5. OSSE designs and results

### 5.1. Background field from model output

In this section a suitable treatment is sought for a background field which is required by VDRAS, but inevitably contains incorrect information. Four experiments (BGR1, BGR1SM, BGR2 and BGR2SM) are designed to explore this topic. These experiments have the same assimilation and forecast cycles as illustrated in Figure 2, but with different backgrounds and first guesses at the beginning of the first cycle (i.e. the cold start). The assimilation period

 

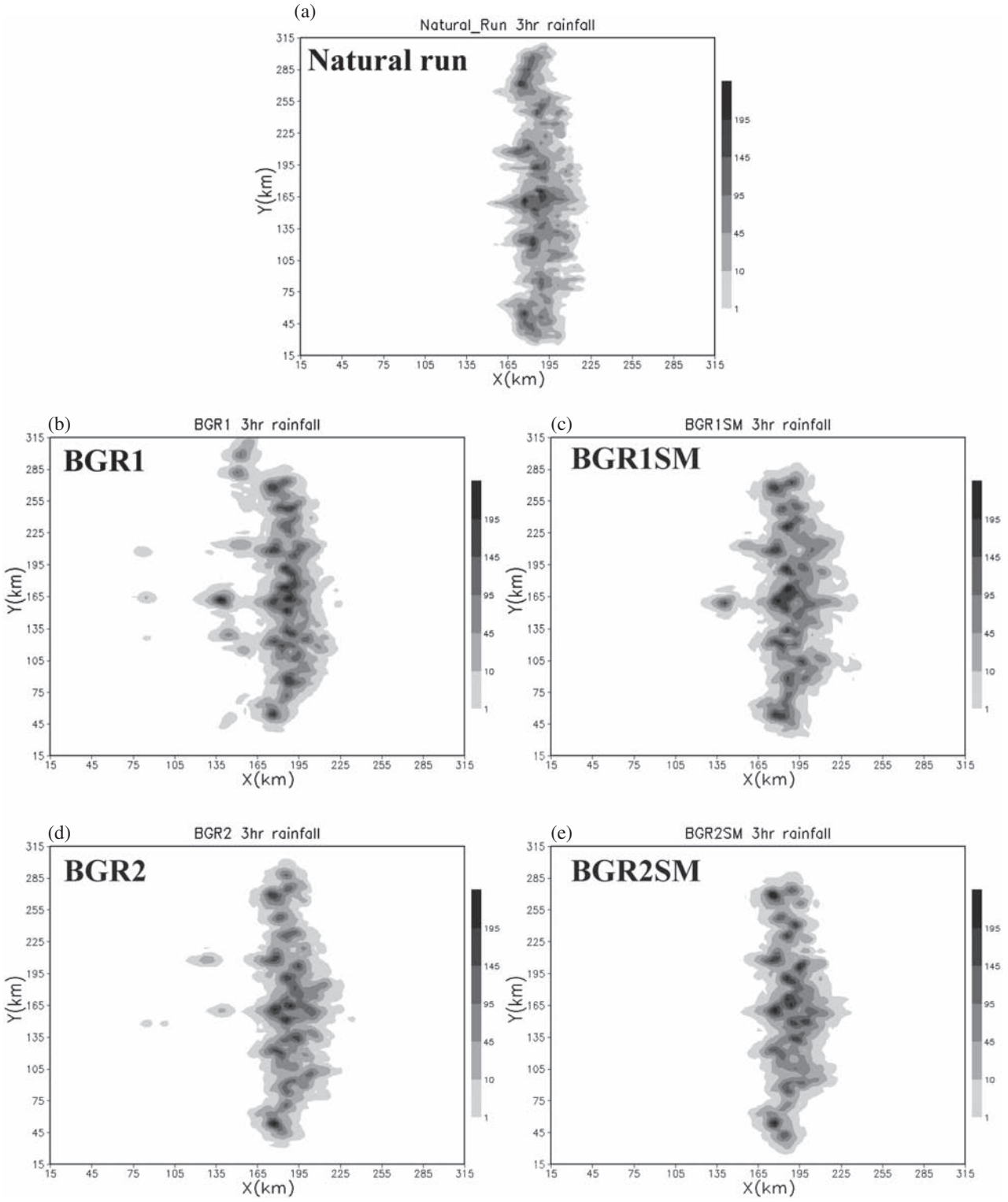


Figure 5. (a) The 3 h accumulated rainfall amount (in mm) from the natural run starting from a simulation time of $T = 10\,800$ s. (b–e) The predicated 3 h accumulated rainfall amount (in mm) from experiments BGR1, BGR1SM, BGR2 and BGR2SM, respectively.

begins at a corresponding simulation time of the natural run at $T = 9300$ s, followed by two 10 min 4DVar cycles. Each 4DVar cycle assimilates both radial wind and reflectivity (computed from the natural run) at three time levels (or volume scans), with an interval of 5 min. The data from radar 1 and radar 2 are assimilated synchronously. The first and last volumes fit the starting and ending times of each cycle. A 5 min forecast is inserted between 4DVar cycles, and provides the background and first guess for the following cycle. A 3 h forecast follows the assimilation period. This is our default assimilation and forecast design for the OSSEs.

The two simulations, R1 and R2, are designed, such that the information provided by these runs represents the

 

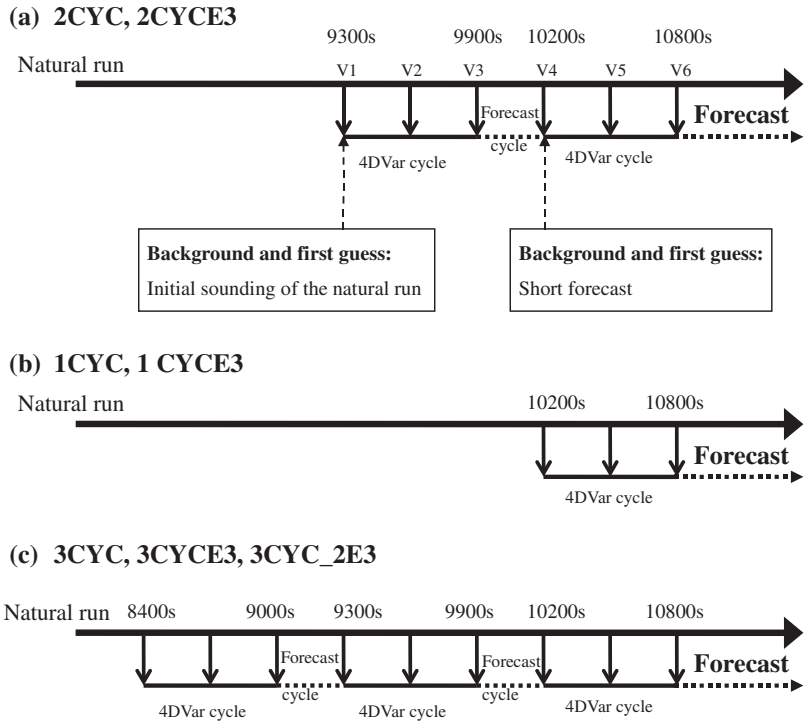


Figure 6. Illustration of VDRAS assimilation and forecast cycles for the experiments (a) 2CYC and 2CYCE3; (b) 1CYC and 1CYCE3; (c) 3CYC, 3CYCE3 and 3CYC_2E3. In these experiments, the background and first guess for the first 4DVar cycle are provided by the initial sounding of the natural run.

Table 1. A list of the experiments.

| Experiment | Assimilation cycles | | | | Description |
|---|---|---|---|---|---|
| | One cycle | Two cycles | Three cycles | Long cycle | |
| BGR1 | | X | | | R1 as background and first guess |
| BGR1SM | | X | | | Smoothed R1 as background and first guess |
| BGR2 | | X | | | R2 as background and first guess |
| BGR2SM | | X | | | Smoothed R2 as background and first guess |
| 1CYC | X | | | | With different numbers of 4DVar cycles, and sounding as background and first guess |
| 2CYC | | X | | | |
| 3CYC | | | X | | |
| 1CYCE3 | X | | | | Same as 1CYC, 2CYC and 3CYC, but lacking radar data below 3° elevation angle |
| 2CYCE3 | | X | | | |
| 3CYCE3 | | | X | | |
| 3CYC_2E3 | | | X | | Same as 3CYCE3, but lacking radar data below 3° elevation angle for the first two cycles |
| L5V_5Z | | | | X | Long assimilation cycle covering five radar volumes |
| L5V_3Z | | | | X | Long assimilation cycle, but assimilating the radial wind for five volumes and reflectivity for the first three volumes |
| L5V_3ZL | | | | X | Same as L5V_3Z, but assimilating the reflectivity for the last three volumes |
| R2CYC | | X | | | Real case studies |
| RL5V_5Z | | | | X | |
| RL5V_3Z | | | | X | |

incorrect background fields from a mesoscale model output. The convection lines initialized in R1 and R2 have the same maximum potential temperature excess, 1.8 K, which is 0.2 K lower than that in the natural run. In addition, the position of the initial temperature perturbation in R1 is deliberately shifted 42 km westward as compared to the natural run, while in R2 the location is unchanged. The background fields for the first 4DVar cycle in experiments BGR1 and BGR1SM are all provided by R1. The difference is that in the latter experiment, the model output (horizontal wind, temperature and water vapour mixing ratio) from R1 is interpolated and smoothed by the Barnes technique (Barnes, 1964). The formulation for the

 

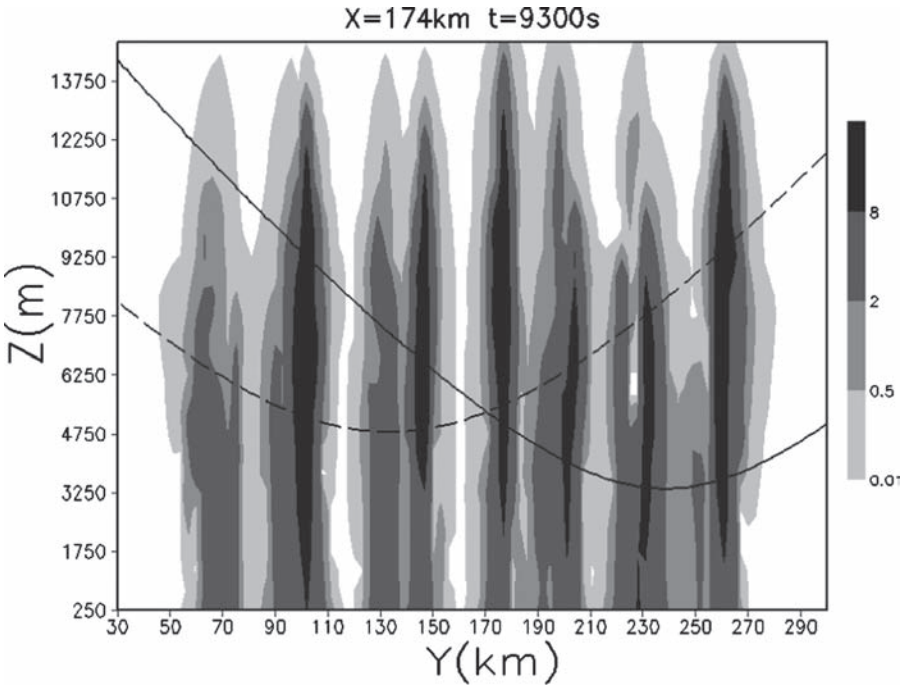


Figure 7. Rainwater mixing ratio (shading) in $g\,kg^{-1}$ in a vertical cross section through $X = 174$ km for the natural run at a simulation time of $T = 9300$ s. The solid and dashed lines in this cross section indicate the height of radar beams with an elevation angle of 3° for radar 1 and radar 2, respectively. The curvature of the Earth is considered when computing the height of the radar beams.

Barnes scheme can be expressed by:

$$\phi_j = \frac{\sum_{i=1}^{N} W_i \phi_i^o}{\sum_{i=1}^{N} W_i} \tag{9}$$

$$W_i = e^{-f(r/R)^2} \tag{10}$$

where $\phi_j$ is the interpolated value of variable $\phi$ at grid point $j$, $\phi_i^o$ denotes the value of variable $\phi$ at selected grid point $i$ with coarse spatial resolution, $W_i$ stands for the weighting coefficient, $r$ represents the distance between the interpolated and the selected grid points, $R$ denotes the radius of influence and $f$ is used to control the smoothness. In this study, $R$ and $f$ are set to 50 km and 2.3, respectively.

The first guess for each cycle is specified to be the same as the background field, except for the rainwater. In VDRAS, the first guess for rainwater is always computed from the observed radar reflectivity through Equation 4. This is because the radar reflectivity is a rather realistic description of the atmosphere. Therefore, by using the reflectivity-converted rainwater in the first guess, the errors such as misplaced or spurious convective systems embedded in the original background field can be corrected effectively. Experiments BGR2 and BGR2SM denote a second set of background field test in which the squall line from R2 is also weaker, but with a relatively correct position when compared against the natural run.

Figure 3(a) and (b) illustrates the horizontal distribution of rainwater produced by R1 and R2 at $T = 9300$ s, respectively. As expected from the experimental designs, the strength of the convective cells in R1 and R2 is weaker than in the natural run. The convective cells in R1 are shifted further westward, while in R2 the location is similar to the natural run (Figure 1(a)).

Figure 3(c) and (d) displays the simulated $u$ component of wind obtained from R1 and R2, while their smoothed counterparts are illustrated in Figure 3(e) and (f), respectively. It can be seen that in the smoothed wind field, the small scale structure of the convective cells has been removed. However, the wind speed outside of the convective cells in Figure 3(e) and (f) remains similar to that in Figure 3(c) and (d), implying that the environmental wind field is generally intact after the smoothing.

A comparison of the ETSs of predicted accumulated rainfall with a threshold of heavy precipitation (i.e. 1.25 mm $(5\,min)^{-1}$) between BGR1 and BGR1SM is exhibited in Figure 4(a). It can be seen that in BGR1SM, a smoothed background indeed improves the rainfall forecast. However, Figure 4(b) reveals that the QPF capability of heavy rain in BGR2SM does not differ from BGR2 significantly. This is because, in BGR2, it is assumed that the location of the squall line in the background field has been rather correctly predicted. It is the initial temperature perturbation that is underestimated, but this discrepancy can be recovered to a certain extent through the radar data assimilation.

The above section discusses the forecast of heavy precipitation. To understand the model's QPF capability for a longer period of time, the spatial distributions of a 3 h accumulated precipitation obtained from the natural run and the four experiments are exhibited in Figure 5. The natural run (Figure 5(a)) shows a clear quasi-linear distribution of precipitation. However, as can be seen in Figure 5(b), because of erroneous convective-scale perturbations from the background field, BGR1 produces spurious scattered rainfall to the west of the squall line. Nevertheless, by a proper smoothing of the background field, the over-prediction of the rainfall can be mitigated in BGR1SM (Figure 5(c)). In Figure 5(d), because the location of the initial convective cells in BGR2 is sufficiently accurate, much less scattered rainfall is predicted if compared

 

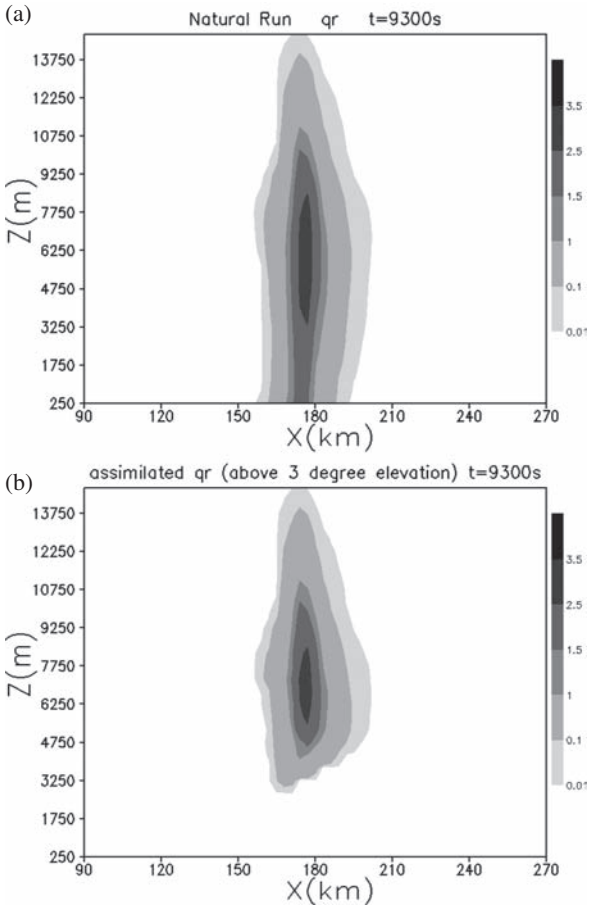


Figure 8. (a) Line-average of the rainwater mixing ratio in $g\,kg^{-1}$ for the natural run at a simulation time of $T = 9300$ s along a vertical cross section. (b) Same as (a), but is the rainwater existing above the observed elevation angle 3° for radar 1 and radar 2.

against BGR1. Figure 5(e) indicates that further suppression of the over-predicted scattered rainfall is achieved in BGR2SM.

### 5.2. Impact of the lack of low-elevation radar observation

In this section, the impact of the absence of low-elevation radar observations on the retrieved analysis fields and rainfall prediction is examined. The background field and first guess for all experiments are provided by a sounding. As illustrated in Figure 6 and Table 1, experiment 2CYC includes two 4DVar cycles with one forecast period inserted in between. The 2CYCE3 is the same as 2CYC, but by assuming that radar reflectivity and radial wind observations are absent below the 3° elevation angle. Experiments 1CYC and 3CYC are similar to 2CYC, but with one and three assimilation cycles, respectively. Similarly, in 1CYCE3 and 3CYCE3, radar data are assumed to be absent below 3°. Experiment 3CYC_2E3 has three assimilation cycles, but the missing of low-elevation radar data only occurs in the first two cycles. Full radar coverage is assumed at the third cycle.

Figure 7 shows a vertical cross section penetrating the areas with active convective cells embedded in the squall line ($X = 174$ km) from the natural run at $T = 9300$ s. The heights of the radar beams from two pseudo-radars with 3° elevation angle are also plotted in Figure 7. The lowest radar beams reach 3250 and 4750 m AGL for radar 1 and radar 2, respectively.

Figure 8(a) reveals the line-averaged (defined as an average of the available radar data along the squall line) rainwater mixing ratio along a vertical cross section from the natural run at $T = 9300$ s, while Figure 8(b) depicts the data availability when low-elevation radar observation is missing. Compared with Figure 8(a), it can be seen that in Figure 8(b), the rainwater (also an indicator of the radar data) gradually decreases downward from the convection core at $Z = 5500$ m, then completely disappears below $Z = 2500$ m.

Figure 9 illustrates the line-averaged temperature perturbation and vertical velocity at $T = 10\,800$ s along a vertical cross section from the natural run, and seven experiments. An examination of Figure 9(a) for the natural run shows a widespread warming area in the upper levels, induced by heating from condensation during the long period of model simulation. The temperature gradient along the horizontal direction is weak. The updraft is $>2.5\,m\,s^{-1}$, located near the warm core at $Z = 6250$ m. The evaporative cooling also produces a cold pool with a depth of 1250 m and a width of 90 km along the surface. Typical upward velocity is induced at the leading edge of the cold pool at $X$ 185 km. Downdrafts about $-0.2\,m\,s^{-1}$ are found behind at $X \sim 175$ km.

In contrast, even with full radar coverage, 1CYC and 2CYC (Figure 9(b) and (c)) show that the retrieved warm area turns out to be much narrower, with a stronger horizontal temperature gradient, leading to downward motions at mid-levels. The retrieved cold pool is also weaker in strength. However, by comparing the experiment 2CYC with 3CYC (Figure 9(c) and (d)), it can be found that as the number of 4DVar cycles increases, the warm area also spreads outward gradually. The strength and width of the cold pool, and the pattern of the mid-level downdrafts, all become more similar to those in the natural run (Figure 9(a)).

When low-elevation observations are absent, Figure 9(e) shows that both the upper-level warm zone and the near-surface cold pool retrieved in 1CYCE3 are dramatically weaker than in 1CYC. The up- and down-drafts at the leading edge of the cold pool are nearly unrecognizable (Figure 9(e)). When more assimilation cycles are involved, the intensity of the upper-level warming zones in 2CYCE3 and 3CYCE3 (Figure 9(f) and (g)) can be improved. However, the recovery of the near-surface cold pool and downdrafts where radar data are absent reveals a different scenario. In 2CYCE3 (Figure 9(f)), one more assimilation cycle is indeed helpful in improving the cold pool intensity. Nevertheless, when three cycles are applied, the cold pool is retrieved incorrectly (Figure 9(g)) with much greater depth and magnitude. This is confirmed by comparing the RMSEs of the retrieved cold pool in these three experiments, as shown in Figure 10. This discrepancy can be explained in that when the rainwater information at lower atmosphere is not assimilated into the model owing to the lack of low-elevation radar data, the model mistakenly suppresses the development of the convection, triggering a reduction of the latent heat release, a decrease of the buoyancy and an over-enhancement of the downdrafts in the mid to low levels (Figure 9(g)). Consequently, the excess amount of rainfall from the upper layers further intensifies the cold pool. However, when the third cycle contains full radar data coverage (3CYC_2E3), Figure 9(h) shows a noticeable improvement of the cold pool structure.

In terms of the accuracy of the retrieved cold pool intensity, when full radar coverage is possible, more assimilation cycles produce better retrievals, as shown by comparing the RMSEs of 1CYC, 2CYC and 3CYC. When low-elevation observations

 

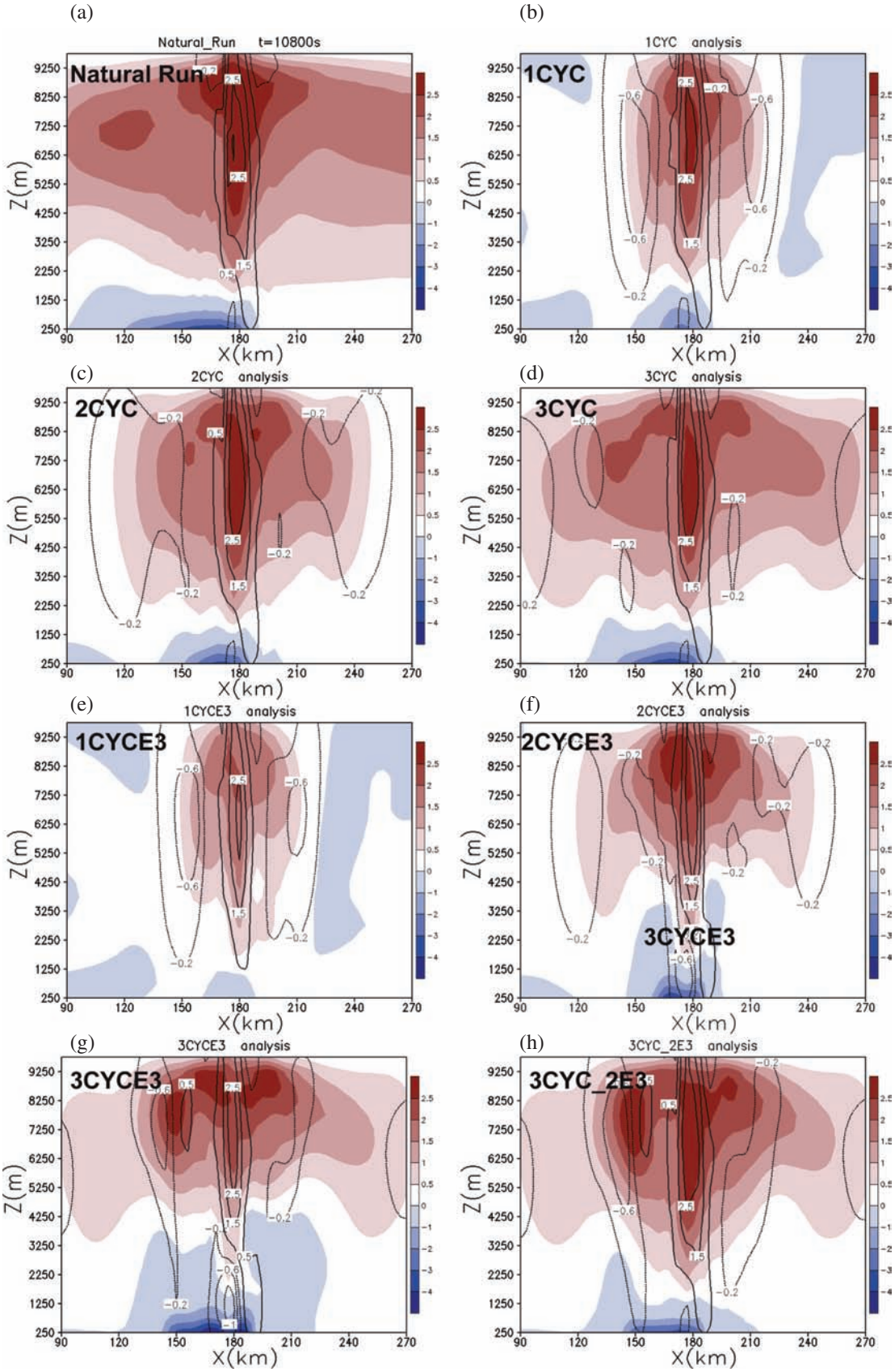


Figure 9. Line-averaged temperature perturbation (coloured in °C) and vertical velocity (contours at 2.5, 1.5, 0.5, −0.2, −0.6 and −1.0 m s$^{-1}$) for (a) the natural run at $T = 10\,800$ s, and experiments: (b) 1CYC, (c) 2CYC, (d) 3CYC, (e) 1CYCE3, (f) 2CYCE3, (g) 3CYCE3 and (h) 3CYC_2E3.

 

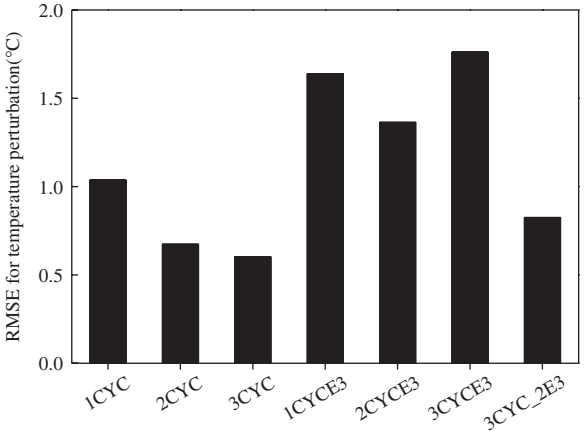


Figure 10. The RMSEs for temperature perturbation (°C) between the natural run and different experimental analyses. The computation is limited within the cold pool area where the reflectivity is larger than 0 dBZ and the height is below 3250 m.

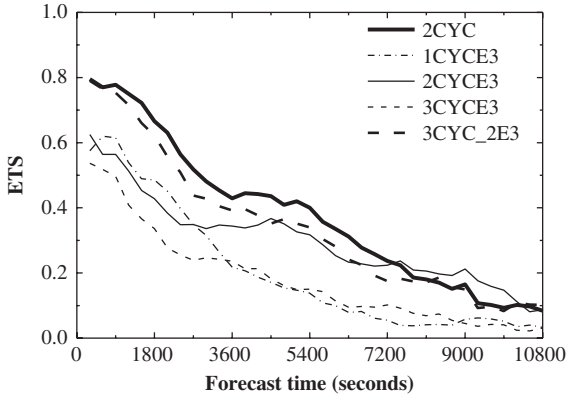


Figure 11. As in Figure 4, but for 2CYC (thick solid line), 1CYCE3 (dash-dotted line), 2CYCE3 (thin solid line), 3CYCE3 (thin dashed line) and 3CYC_2E3 (thick dashed line).

are not available, more assimilation cycles do not always imply better retrievals. Instead, two assimilation cycles appear to be an optimal selection.

Regarding the rainfall prediction, under the condition of full radar coverage, the forecast abilities of 1CYC, 2CYC and 3CYC do not differ from each other dramatically, although 3CYC does show slight improvement of the rainfall prediction at a later stage (>75 min) of the model integration (not shown). However, when low-elevation data coverage is incomplete, Figure 11 indicates that experiment 2CYCE3 outperforms 1CYCE3 and 3CYCE3 after $T = 2700$ s. Experiment 3CYC_2E3 offers additional improvement of the rainfall forecast as long as one of the cycles contains low-elevation observations.

### 5.3. Non-linear microphysical process and assimilation strategy

The study by Fabry and Sun (2010, hereafter FS10) indicates that the model non-linearity could force the minimization procedure towards a wrong direction. Because the 4DVar algorithm seeks an optimal model initial condition, this implies that more data assimilations could lead the model initial condition to deviate from the true state. Furthermore, they also demonstrate that the non-linearity induced by microphysical parameters is greater than that from the wind velocities. Thus, different assimilation strategies should be considered for different model state variables. For highly non-linear variables, a shorter assimilation cycle is suggested.

This section is devoted to test the above-mentioned argument. As displayed in Figure 12, in all experiments, radial winds are assimilated over the entire 20 min long 4DVar cycle using radar data from five volume scans with a 5 min separation. In L5V_5Z, L5V_3Z and L5V_3ZL, the radar reflectivity values (a parameter directly linked to microphysical variables) from all five volume scans, the first three and the last three scans are assimilated into the model, respectively. Figure 13 exhibits the variation of the rainwater term in the cost function (Equation 1) based on radar 1 observations. The computation is conducted after the radar data in volumes V2–V6 are sequentially assimilated from time levels T2–T6. A smaller (larger) initial value of the rainwater cost function implies a good (bad)

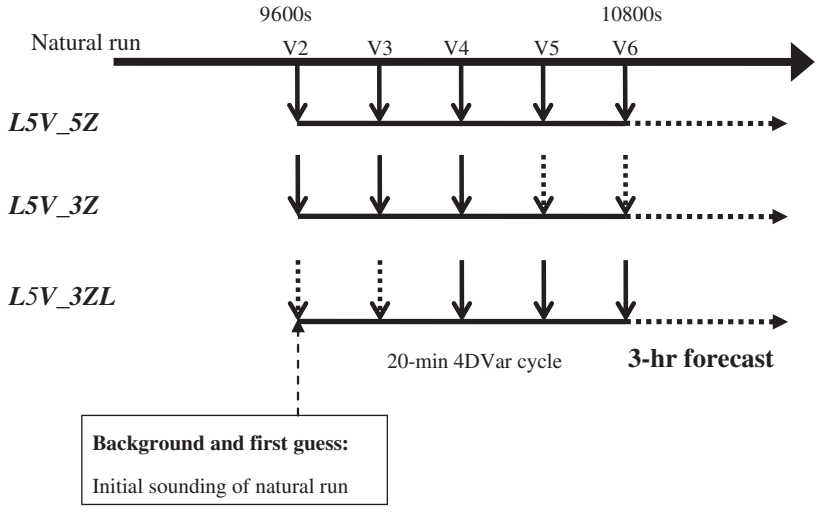


Figure 12. Illustration of VDRAS assimilation and forecast cycles for the experiments: L5V_5Z, L5V_3Z and L5V_3ZL. The 20 min long 4DVar cycle starts from a simulation time of $T = 9600$ s for the natural run, corresponding to the assimilation time for the radar volume V2 in 2CYC (Figure 6(a)). The dashed down-pointing arrows indicate assimilation of only the radial wind.

 

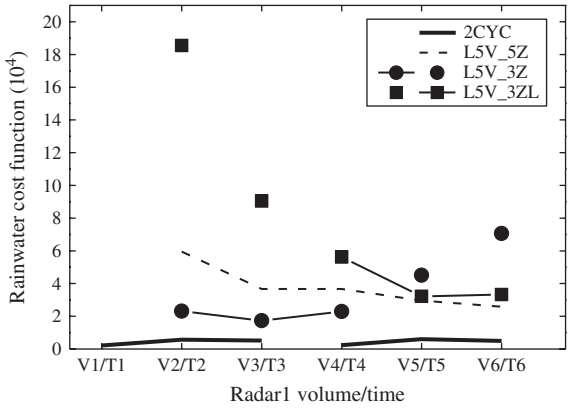


Figure 13. Change of the rainwater term in the cost function shown in Equation 1 with respect to the assimilated radar volumes/times V1–V6/T1–T6 for radar 1 for experiments 2CYC (solid line), L5V_5Z (dashed line), L5V_3Z (—●—) and L5V_3ZL (—■—). Solid circles (●) and solid squares (■) represent the un-assimilated volumes for L5V_3Z and L5V_3ZL, respectively. T1–T6 stand for assimilation time from $T = 9300$ to 10 800 s with a time interval of 300 s.

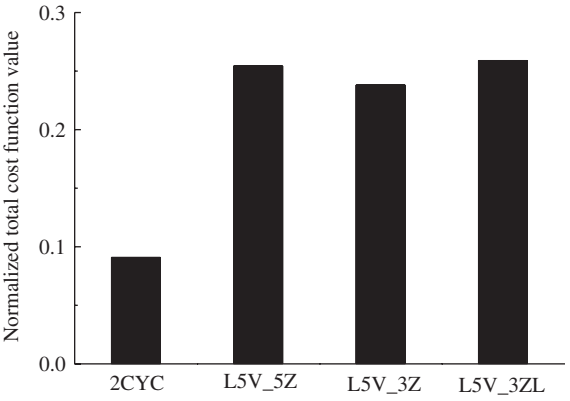


Figure 14. Normalized total cost function values for the experiments 2CYC, L5V_5Z, L5V_3Z and L5V_3ZL.

minimization procedure, leading to a better (worse) fit of the model and the observations. The cost function magnitude from 2CYC (Figure 6(a)) is also displayed in Figure 13 as a reference.

The behaviour of the cost function exhibited by L5V_3Z indicates that by assimilating the radar reflectivity only at the first three time levels (representing an earlier stage or 'linear regime' in FS10), a small initial value is obtained, meaning that the minimization procedure is properly executed. When reflectivity is assimilated in all five time levels (T2–T6), the initial value of the cost function from L5V_5Z becomes larger, indicating that the model non-linearity has begun to drive the model away from the true initial state. The worst case occurs when the reflectivity is assimilated only at the last three time levels (representing a later stage or 'contradictory regime' in FS10); the cost function from L5V_3ZL reaches the largest initial value. This suggests that because of the non-linearity, the minimization algorithm has driven the model towards a wrong direction, and is unable to find an optimal initial condition that could minimize the misfit between the following model simulations and the observations within the assimilation window. Finally, instead of using a single long assimilation cycle, experiment 2CYC uses two short cycles, in which both radial wind and reflectivity are assimilated. The

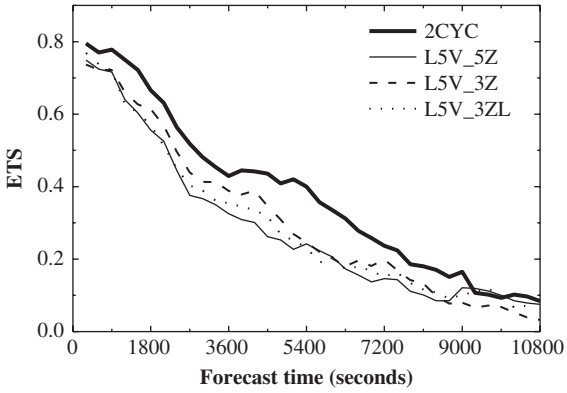


Figure 15. As in Figure 4, but for 2CYC (thick solid line), L5V_5Z (thin solid line), L5V_3Z (dashed line), and L5V_3Z (dotted line).

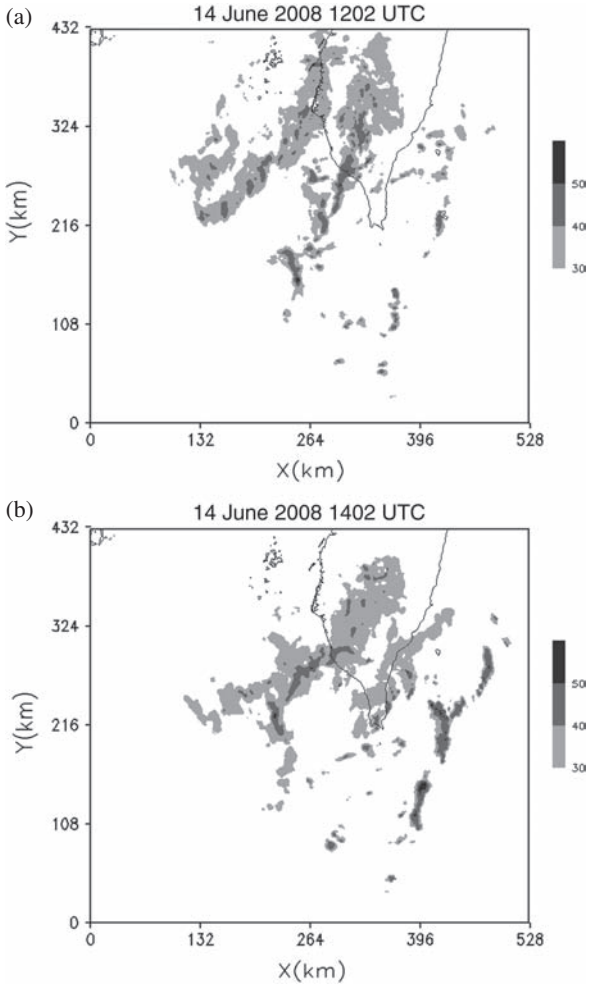


Figure 16. The composite maximum radar reflectivity (grey shading in dBZ with 10 dBZ interval) on 14 June 2008 for IOP #8 at (a) 1202 and (b) 1402 UTC. The solid line represents the coast line of Taiwan.

short assimilation cycle helps to keep the model trajectory in the nearly linear regime, thus the initial magnitude of the cost function in 2CYC is substantially smaller than that from all other experiments, as depicted in Figure 13.

Experiments were conducted (not shown) in which reflectivity are assimilated at all five volumes, but radial wind is assimilated at selected time steps (e.g. first three or last three

 

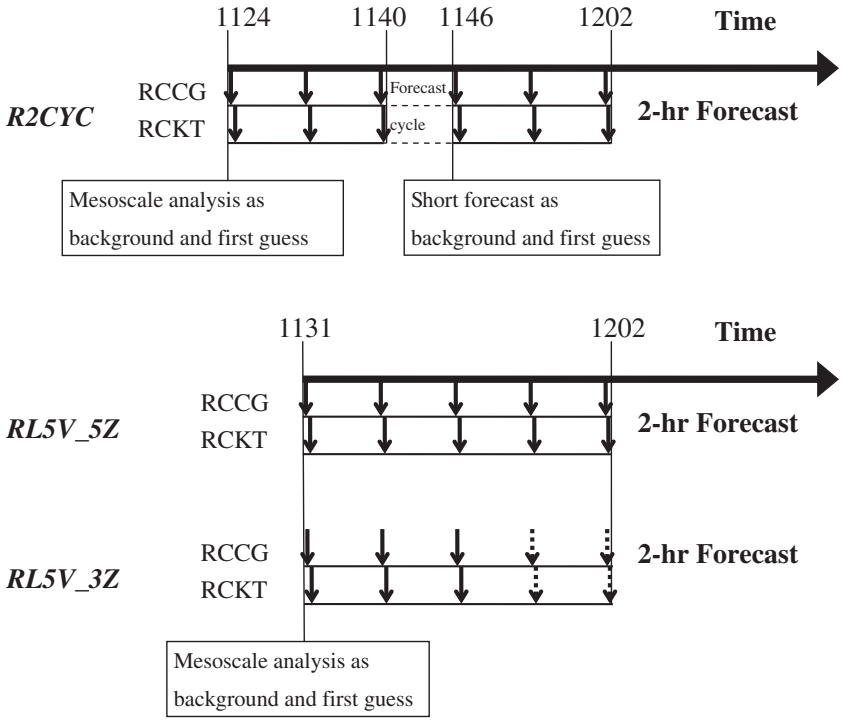


Figure 17. Illustration of assimilation and forecast cycles for the experiments R2CYC, RL5V_5Z and RL5V_3Z. The numbers at the top of the diagram indicate the starting and finishing times (UTC) of each 4DVAR cycle. RCKT and RCCG stand for the two radars, respectively.

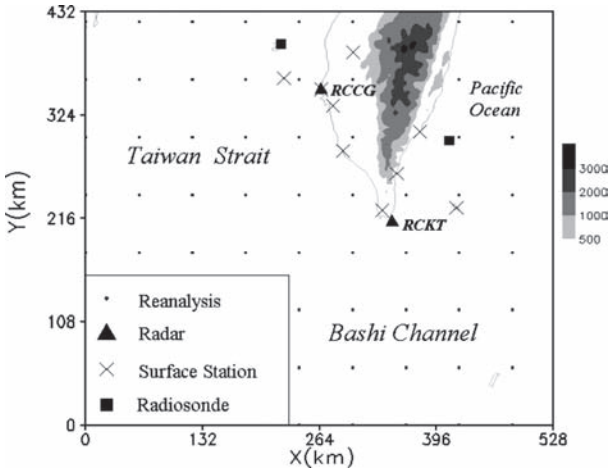


Figure 18. Locations of data collected from different sources for SoWMEX IOP #8. The small dots indicate selected data points from the ECMWF reanalysis, triangles stand for radars, crosses represent surface mesonet stations, and solid circles denote radiosondes. This is also the experimental domain for the VDRAS.

volumes only). It is found that the non-linearity phenomenon as shown in Figure 13 does not appear. Assimilating radial wind into the model at different stages through 4DVar technique still gives a similar 'optimal' initial condition. This result implies that the non-linearity induced by reflectivity dominates that from radial wind, and is consistent with the finding by FS10.

This section focuses on the minimization of the total cost function and compares the QPF scores produced by all experiments. Figure 14 shows that the best minimization is achieved by 2CYC. Owing to the non-linearity, using a single long assimilation cycle gives inferior minimization, resulting in larger cost function values in experiments L5V_5Z, L5V_3Z and L5V_3ZL. Nevertheless, L5V_3Z produces a slightly better result, because the reflectivity is assimilated only at an early stage when the model is still in its linear regime. As expected, Figure 15 indicates that when two short cycles are employed, 2CYC gives the best QPF score, followed by L5V_3Z. These experimental results clearly point out that when designing a proper assimilation algorithm, the non-linearities and their different growing speeds inherited in various model state variables have to be considered separately.

## 6. Microphysical non-linearity and assimilation strategy – a real case study of 2008 SoWMEX IOP #8

### 6.1. A brief introduction of 2008 SoWMEX IOP #8

To further investigate the influence of the model non-linearity under a realistic framework, a real case from SoWMEX IOP #8, a field experiment conducted in 2008 in Taiwan, was selected. This particular IOP stands for a pre-frontal squall line, whose influence on the precipitation in Taiwan lasts for about 2 days. Thus, the entire observation period started at 0000 UTC 14 June, and ended at 0000 UTC 16 June. Figure 16 illustrates the distributions of the composite maximum radar reflectivity on 14 June at 1202 and 1402 UTC, observed by two CWB operational S-band Doppler radars (RCCG and RCKT; see Figure 18 for their geographic locations). The convection systems, elongated in an NE–SW direction and moving towards the east and southeast, produced heavy precipitation over southern Taiwan.

It should be pointed out that the other limitations discussed in Sections 5.1 and 5.2, such as an imperfect background field and the absence of low-level radar observations, are also present in this real case study. Thus, this section stands for an overall verification of the conclusions obtained from the OSSEs.

 

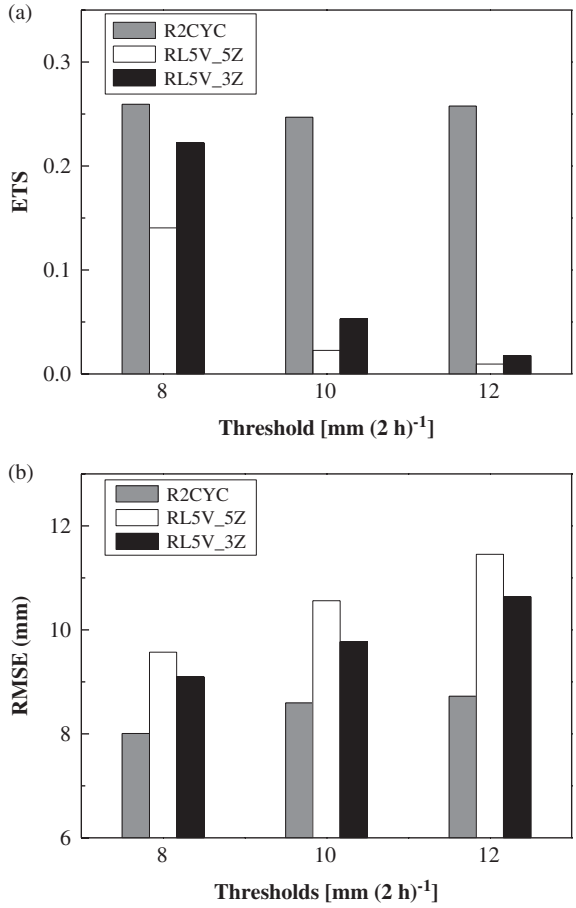


Figure 19. (a) ETSs and (b) RMSEs of the forecast 2 h rainfall accumulations over the island of Taiwan from the experiments R2CYC, RL5V_5Z and RL5V_3Z at different precipitation thresholds.

### 6.2. Experimental designs and results

Experiments R2CYC, RL5V_5Z and RL5V_3Z are equivalent to their OSSE counterparts 2CYC, L5V_5Z and L5V_3Z, respectively. As shown in Figure 17, the assimilation period for R2CYC includes two 16 min 4DVar cycles, with a short 6 min forecast period inserted in between. Experiments RL5V_5Z and RL5V_3Z contain a single 31 min 4DVar cycle. Five radar data volumes including radial winds and reflectivity are assimilated in RL5V_5Z. Experiment RL5V_3Z is similar to RL5V_5Z, but the reflectivity is assimilated only at the first three time levels. The assimilation for all experiments ends at 1202 UTC, followed by a 2 h forecast.

Figure 18 illustrates the locations of two CWB S-band Doppler radars (RCCG and RCKT), nine surface stations and two radiosonde stations. The re-analysis data from the European Centre for Medium-Range Weather Forecasts (ECMWF) are also adopted to fill in the data-void region over the ocean. Except for the radar measurements, the other *in situ* observations and the ECMWF reanalysis data are combined to establish a mesoscale background field using the Barnes interpolation scheme discussed in Section 5.1. Radar data are carefully corrected (i.e. velocity unfolding, ground clutter removal) before being ingested into VDRAS, which has a domain size of $528 \times 432 \times 15\,\mathrm{km}^3$, with the horizontal (vertical) grid space set to 2.0 km (500 m). It should be pointed out that owing to the PPI scanning strategy of the radars and blockage of complex topography (with a peak height reaching almost 4000 m), the lack of low-level radar observations is a frequent problem in Taiwan. The deployment of low-power/short-range radars to fill in data-void region should be a future consideration.

By comparing the ETSs and RMSEs under different rainfall thresholds (8, 10, 12 mm $(2\,\mathrm{h})^{-1}$), Figure 19 illustrates that the QPF ability of R2CYC is indeed superior to that of RL5V_3Z. As expected, RL5V_5Z has the lowest scores, apparently due to the assimilation of the reflectivity over a single long cycle. The finding from this set of experiments is in agreement with the conclusions obtained from OSSEs presented in previous sections.

## 7. Summary and conclusions

In this study, a series of OSSEs and a real case experiment are conducted to investigate the influence of background error, the lack of low-elevation radar observation and microphysical non-linearity on the application of Doppler radar data assimilation technique to QPF. The major conclusions are summarized as follows.

1. When a mesoscale model output, which inevitably contains incorrect information, is used as the background field for the data assimilation system, a smoothing procedure is suggested to remove the fine structure embedded in the model-simulated convective systems. This procedure can help to suppress the growth of spurious precipitations, and improve the model's QPF ability.
2. Lack of low-elevation radar observations for a weather system has a negative impact on the retrieval of the near-surface cold pool. When low-elevation data are missing, the number of assimilation cycles needs to be carefully selected. It is found that the increase of the assimilation cycles does not necessarily improve the cold pool retrieval. The OSSE tests in this study suggest that two assimilation cycles are optimal for the recovery of low-level cold pool, and produce a better rainfall forecast.
3. Both OSSEs and the real case experiment reveal that the non-linearity developed from the microphysical variables could worsen the data assimilation and minimization procedure and prevent the algorithm from finding the optimal model initial state. Thus, the radar reflectivity should be assimilated only within a short cycle and at the early stage of the model integration. The strategy of using multiple short cycles enables the minimization algorithm to proceed towards the right direction and to approach the correct initial condition. As a result, a model with better QPF ability can be achieved.

## Acknowledgements

The authors are grateful to the SoWMEX project team. The two anonymous reviewers are also highly appreciated for their valuable comments that helped to improve this paper. This research was supported by the National Science Council of Taiwan, under NSC101-2625-M-008-003 and NSC101-2119-M-008-019.

## References

Barnes S. 1964. A technique for maximizing details in numerical map analysis. *J. Appl. Meteorol.* **3**: 395–409.

 

Chung K-S, Zawadzki I, Yau MK, Fillion L. 2009. Short term forecasting of a midlatitude convective storm by the assimilation of single–Doppler radar observations. *Mon. Weather Rev.* **137**: 4115–4135.

Crook NA. 1994. Numerical simulations initialized with radar-derived winds. Part I: Simulated data experiments. *Mon. Weather Rev.* **122**: 1189–1203.

Crook NA, Sun J. 2002. Assimilating radar, surface, and profiler data for the Sydney 2000 forecast demonstration project. *J. Atmos. Oceanic Tech.* **19**: 888–898.

Crook NA, Tuttle JD. 1994. Numerical simulations initialized with radar-derived winds. Part II: Forecasts of three gust-front cases. *Mon. Weather Rev.* **122**: 1204–1217.

Fabry F, Sun J. 2010. For how long should what data be assimilated for the mesoscale forecasting of convection and why? Part I: On the propagation of initial condition errors and their implications for data assimilation. *Mon. Weather Rev.* **138**: 242–255.

Gal-Chen T. 1978. A method for the initialization of the anelastic equations: implications for matching models with observations. *Mon. Weather Rev.* **106**: 587–606.

Gao J, Stensrud DJ. 2012. Assimilation of reflectivity data in a convective-scale, cycled 3DVAR framework with hydrometeor classification. *J. Atmos. Sci.* **69**: 1054–1065.

Hu M, Xue M, Gao J, Brewster K. 2006a. 3DVAR and cloud analysis with WSR-88D level-II data for the prediction of the Fort Worth, Texas, tornadic thunderstorms. Part I: Cloud analysis and its impact. *Mon. Weather Rev.* **134**: 675–698.

Hu M, Xue M, Gao J, Brewster K. 2006b. 3DVAR and cloud analysis with WSR-88D level-II data for the prediction of the Fort Worth, Texas, tornadic thunderstorms. Part II: Impact of radial velocity analysis via 3DVAR. *Mon. Weather Rev.* **134**: 699–721.

Jung Y, Zhang G, Xue M. 2008a. Assimilation of simulated polarimetric radar data for a convective storm using the ensemble Kalman filter. Part I: Observation operators for reflectivity and polarimetric variables. *Mon. Weather Rev.* **136**: 2228–2245.

Jung Y, Xue M, Zhang G, Straka JM. 2008b. Assimilation of simulated polarimetric radar data for a convective storm using the ensemble Kalman filter. Part II: Impact of polarimetric data on storm analysis. *Mon. Weather Rev.* **136**: 2246–2260.

Klemp JB, Wilhelmson RB. 1978. The simulation of three- dimensional convective storm dynamics. *J. Atmos. Sci.* **35**: 1070–1096.

Li Y, Wang X, Xue M. 2012. Assimilation of radar radial velocity data with the WRF ensemble-3DVAR hybrid system for the prediction of hurricane Ike (2008). *Mon. Weather Rev.* **140**: 3507–3524.

Lin Y, Peter SR, Kenneth WJ. 1993. Initialization of a modeled convective storm using Doppler radar–derived fields. *Mon. Weather Rev.* **121**: 2757–2775.

Pan X, Tian X, Li X, Xie Z, Shao A, Lu C. 2012. Assimilating Doppler radar radial velocity and reflectivity observations in the weather research and forecasting model by a proper orthogonal-decomposition-based ensemble, three-dimensional variational assimilation method. *J. Geophys. Res.* **117**: D17113, DOI: 10.1029/2012JD017684.

Rogers E, Black TL, Deaver DG, DiMego GJ, Zhao Q, Baldwin M, Junker NW, Lin Y. 1996. Changes to the operational "early" Eta analysis/forecast system at the National Centers for Environmental Prediction. *Weather Forecast* **11**: 391–412.

Schaefer JT. 1990. The critical success index as an indicator of warning skill. *Weather Forecast* **5**: 570–575.

Snyder C, Zhang F. 2003. Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. *Mon. Weather Rev.* **131**: 1663–1677.

Sun J. 2005. Initialization and numerical forecasting of a supercell storm observed during STEPS. *Mon. Weather Rev.* **133**: 793–813.

Sun J, Chen M, Wang Y. 2010. Frequent-updating analysis system based on radar, surface, and mesoscale model data for the Beijing 2008 Forecast Demonstration Project. *Weather Forecast* **25**: 1715–1735.

Sun J, Crook NA. 1997. Dynamical and microphysical retrieval from Doppler radar observations using a cloud model and its adjoint. Part I: Model development and simulated data experiments. *J. Atmos. Sci.* **54**: 1642–1661.

Sun J, Crook NA. 2001. Real-time low-level wind and temperature analysis using WSR-88D data. *Weather Forecast* **16**: 117–132.

Sun J, Wang H. 2013. Radar data assimilation with WRF 4D-Var: Part II. Comparison with 3D-Var for a squall line over the U.S. Great Plains. *Mon. Weather Rev.* **141**: 2245–2264.

Sun J, Zhang Y. 2008. Analysis and prediction of a squall line observed during IHOP using multiple WSR-88D observations. *Mon. Weather Rev.* **136**: 2364–2388.

Tai S-L, Liou Y-C, Sun J, Chang S-F, Kuo M-C. 2011. Precipitation forecasting using Doppler radar data, a cloud model with adjoint, and the weather research and forecasting model: real case studies during SoWMEX in Taiwan. *Weather Forecast* **26**: 975–992.

Takuya K, Kuroda T, Seko H, Saito K. 2011. A cloud-resolving 4DVAR assimilation experiment for a local heavy rainfall event in the Tokyo metropolitan area. *Mon. Weather Rev.* **139**: 1911–1931.

Tong M, Xue M. 2005. Ensemble Kalman filter assimilation of Doppler radar data with a compressible nonhydrostatic model: OSS experiments. *Mon. Weather Rev.* **133**: 1789–1807.

Wang H, Sun J, Zhang X, Huang X-Y, Auligne T. 2013. Radar data assimilation with WRF 4D-Var: Part I. System development and preliminary testing. *Mon. Weather Rev.* **141**: 2224–2244.

Weisman ML, Klemp JB. 1982. The dependence of numerically simulated convective storms on vertical wind shear and buoyancy. *Mon. Weather Rev.* **110**: 504–520.

Weisman ML, Rotunno R. 2004. "Theory for Strong Long-Lived Squall Lines" revisited. *J. Atmos. Sci.* **61**: 361–382.

Weygandt SS, Shapiro A, Droegemeier KK. 2002. Retrieval of model initial fields from single-Doppler observations of a supercell thunderstorm. Part II: Thermodynamic retrieval and numerical prediction. *Mon. Weather Rev.* **130**: 454–476.

Xiao Q, Kuo Y-H, Sun J, Lee W-C, Lim E, Guo Y, Barker DM. 2005. Assimilation of Doppler radar observations with a regional 3DVAR system: Impact of Doppler velocities on forecasts of a heavy rainfall case. *J. Appl. Meteorol.* **44**: 768–788.

Xiao Q, Sun J. 2007. Multiple radar data assimilation and short-range QPF of a squall line observed during IHOP_2002. *Mon. Weather Rev.* **135**: 3381–3404.

Xue M, Tong M, Droegemeier KK. 2006. An OSSE framework based on the ensemble square root Kalman filter for evaluating impact of data from radar networks on thunderstorm analysis and forecast. *J. Atmos. Oceanic Tech.* **23**: 46–66.

Zhao Q, Cook J, Xu Q, Harasti PR. 2006. Using radar wind observations to improve mesoscale numerical weather prediction. *Weather Forecast* **21**: 502–522.

 