JP1J.6

IMPACT OF RADAR RAINFALL DATA ASSIMILATION ON SHORT-RANGE QUANTITATIVE PRECIPITATION FORECASTS USING FOUR-DIMENSIONAL VARIATIONAL ANALYSIS TECHNIQUE

Linus H.Y. Yeung[*]*, Philip K.Y. Chan and Edwin S.T. Lai*
Hong Kong Observatory, Hong Kong, China

1. INTRODUCTION

In Hong Kong, intense precipitation associated with the summer monsoon, monsoon troughs and tropical cyclones poses a major weather threat in the warm season. For short-range quantitative precipitation forecasts (QPF), the Operational Regional Spectral Model (ORSM) at the Hong Kong Observatory currently relies heavily on radar-based rainfall data ingested through a Physical Initialization (PI) process (Matsumura 1995). Together with a "pre-run" procedure (to be explained in more details in Section 3.1), the model so initialized could be regarded as a kind of "warm" start, in the sense that model spin up could be improved.

In order to have a diabatic or "hot" start with respect to precipitation forecast, a better data assimilation scheme with emphasis on radar rainfall data and balanced fields (both dynamic and thermodynamic) is essential. Previous studies have shown that radar-rainfall assimilated by 4DVAR technique could lead to improved precipitation forecasts up to 18 hours (Tsuyuki 2002). To explore the feasibility of operating a 4DVAR data assimilation system (DAS) for Hong Kong, the Meso 4DVAR (JMA 2002) was adapted from the Japan Meteorological Agency. Using the parallel (MPI) version of ORSM, denoted as MPI RSM, numerical simulations were conducted to evaluate the effectiveness of the 4DVAR DAS in the assimilation of precipitation data and the impact on short range QPF.

In Section 2, the design of the impact study and the data sets used are documented. The methodology, including various analysis/modeling/ verification tools, employed for this study is covered in Section 3. Objective forecast verification results are presented in Section 4. Section 5 summarizes the findings and conclusions.

2. DATA SETS FOR IMPACT STUDY

The impact of radar rainfall data and the 4DVAR analysis technique on short-range QPF was assessed by reviewing a collection of heavy rain cases in 2005, based on MPI-RSM short-range QPF performance with and without 4DVAR.

A total of 12 heavy rain cases (Table I) in the rain season of 2005 were included in the QPF verification. For each case, QPF data were obtained by executing a series of 3-hourly analysis-forecast cycle runs. To distinguish from the routine MPI RSM cycle ("Routine Cycle" for short), model runs based on 4DVAR analysis would be referred to as "4DVAR Cycle". The 4DVAR Cycle had to be cold-started from the forecast fields of Routine Cycle. A consequence was that the forecast skills of the initial one or two model runs were significantly poorer. For the 4DVAR Cycle to take full effect, a finite lead time is required. From experience, about 15 hours (i.e. 5 model runs) would be necessary. But to ensure a sufficient sample size, all model runs, including those being cold-started, were included in the QPF verification. In theory, performance of 4DVAR Cycle in operational mode should be better without the cold-start disadvantage.

A total of 52 model runs were performed for the cases studied, each producing 42 one-hour accumulated rainfall forecasts. All such forecasts were included in the QPF verification. The corresponding model runs from Routine Cycle served as the control.

Radar rainfall information observed during the verification periods were taken as the "ground truth". Within the 256-km range of the HKO Doppler weather radar, hourly rainfall distribution was estimated at about 1 km resolution from radar reflectivity data taken

Rain Case	Approx. Start Time	Intensity Class	Mechanism
(1)	2005-05-05 18 UTC	Green	Weak cold front
(2)	2005-05-08 02 UTC	Amber	Monsoon disturbance
(3)	2005-05-09 04 UTC	Amber	Monsoon trough
(4)	2005-05-09 18 UTC	Amber	Monsoon trough
(5)	2005-05-10 03 UTC	Amber	Monsoon trough
(6)	2005-05-18 12 UTC	Amber	Monsoon trough
(7)	2005-05-27 10 UTC	Amber	Monsoon trough
(8)	2005-06-15 06 UTC	Amber	Monsoon trough
(9)	2005-06-21 00 UTC	Amber	Monsoon disturbance
(10)	2005-06-22 22 UTC	Green	Monsoon disturbance
(11)	2005-06-23 16 UTC	Amber	Monsoon trough
(12)	2005-06-24 02 UTC	Red	Monsoon trough

Table I - Heavy rain cases for impact study. All cases involve organized and wide-spread rainstorms affecting Hong Kong. Rainstorm intensity is divided into 4 classes, namely Green, Amber, Red and Black for peak 1-hour rain-gauge accumulated rainfall in the range 20-30 mm/h, 30-50 mm/h, 50-70 mm/h and ≥ 70 mm/h respectively.

[*] *Corresponding author address*: Linus H.Y. Yeung, Hong Kong Observatory, 134A, Nathan Road, Kowloon, Hong Kong, China; e-mail: lhyyeung@hko.gov.hk

at the 3-km level. The estimation was done dynamically with the Z-R relation calibrated in real-time against data from a dense rain-gauge network over Hong Kong.

3. METHODOLOGY

3.1. *MPI RSM*

The MPI RSM is a parallel version of the ORSM based on the Message Passing Interface (MPI). Both versions originated from the Regional Spectral Model of JMA (JMA 2002). Model characteristics and configurations are summarized in Table II. In trial operation at HKO, MPI RSM runs with a one-way two-level nesting strategy. The horizontal resolutions of the inner and outer domains are 20 and 60 km respectively. The external boundary conditions are prepared from the JMA Global Spectral Model data. For the present impact study, only the 20-km model was run.

Identical to ORSM operation, MPI RSM routinely uses the multivariate 3-Dimenional Optimal Interpolation (3DOI) analysis scheme for assimilating conventional observations (see Table II). Rainfall observations are nudged into the forecast model during start up using the PI technique (Matsumura 1995). During PI, thermodynamic fields including temperature, moisture and heating profile (between lifting condensation level and cloud top inferred from the cloud top temperature) over observed rain areas are adjusted through a number of reverse physical processes. The initial conditions needed to start a short-range model integration are ultimately prepared through a 3-hour pre-run process (JMA 2002) prior to analysis time. Forecasts based on 3DOI and PI are hereafter referred to as the "routine" forecasts and act as the control in the present impact study.

3.2. *MPI-RSM 4DVAR*

The 4DVAR DAS employed in the present simulation experiment originated from the Meso 4DVAR of JMA (JMA 2002). In essence, Meso 4DVAR adopts an incremental approach and calculates analysis increments (unbalanced winds, virtual temperature, surface pressure and specific humidity) with an inner-loop model at 20 km resolution (half resolution of the mesoscale forecast model). Whereas the backward integration in the 4DVAR inner loop is done by an adjoint model with reduced physics (i.e. only grid-scale condensation, moist convective adjustment, simplified vertical diffusion and simplified longwave radiation are kept) in a way analogous to other 4DVAR DAS, the forward integration is performed by the full-physics nonlinear model. With such a design, the precipitation term in the observation cost function is calculated more accurately than those relying on linearized model. Another characteristic of Meso 4DVAR is that the variance σ^2 for the precipitation cost function is

	Outer domain	Inner Domain
Model characteristics	Hydrostatic primitive equation model, hybrid σ-P terrain-following coordinates, grid-scale condensation, Arakawa-Schubert cumulus scheme, moist convective adjustment, Mellor-Yamada level-2 non-local PBL scheme.	
Domain coverage	9S-59N, 65E-152E	10N-35N, 100E-128E
Horizontal resolution	60 km	20 km
Vertical levels	40	40
Model top	10 hPa	10 hPa
Initial conditions	60-km analysis	20-km analysis
Boundary conditions	JMA GSM forecast data	60-km MPI RSM forecast data
Analysis Scheme	Single 3DOI+PI	3 hourly pre-runs with 3DOI+PI
Data ingested	SYNOP, BUOY, SHIP, AWS, AMDAR, SATOB, SATEM, ATOVS, wind profiler, radar rainfall	
Analysis time	00, 06, 12, 18 UTC	00, 03, 06, 09, 12, 15, 18, 21 UTC
Forecast range	72 hours	42 hours

Table II – Characteristics and configuration of MPI RSM.

dynamical and dependent on observed rainfall (y_o) and model rainfall (y) as follows:

$$\sigma = \begin{cases} \sigma_o & \text{if} \quad y \le y_o \\ 3\sigma_o & \text{if} \quad y > y_o \end{cases}, \qquad \sigma_o = \max\{1 \text{ mm/h}, y_o\}$$

In other words, over-predicted rainfall will be penalized 3 times heavier than under-prediction. The minimum penalty σ_o is set to either 1 mm/hr or y_o, whichever is the higher. Such a construction is to cope with the exponential distribution of observed precipitation with respect to a given model precipitation value (Koizumi 2003). Together, these two important characteristics make Meso 4DVAR unique and effective in the direct assimilation of precipitation data.

To make Meso 4DVAR run more efficiently on limited computing resource, the resolution of the adjoint model, and hence the analysis increment, has been changed from 20 km to 40 km at HKO. While such a reduction in analysis resolution may sacrifice some capability to capture detailed mesoscale features in highly variable fields such as convective precipitation, the doubling of grid spacing not only reduced the total grid size but also the number of time steps in the assimilation window. Since Meso 4DVAR needs to save all intermediate basic fields during forward integration over the assimilation window, reduction in the number of time steps leads to a drastic saving in computer memory. As will be seen from the affirmative QPF verification results in Section 4, the sacrifice of analysis resolution is justified with respect to forecast performance gain. For easy reference thereafter, we denote the modified version at HKO as MPI-RSM 4DVAR.

3.3. *Rainfall Data Preparation*

Radar rainfall data were prepared as discussed in Section 2. Outside the 256-km radar range, GOES-9 infrared brightness temperature data were used to estimate rainfall rate by a simple regression approach.

Fig. 1 Verification grid of size 16 by 16, centred at Hong Kong. Each grid box is about 28 × 28 km², with a total verification area of about 450 × 450 km².

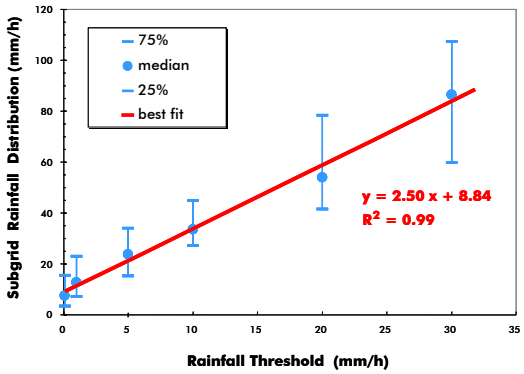


Fig. 2 Sub-grid rainfall distribution against verification threshold values. For simplicity, only the middle 50 % distribution is shown in blue. Also shown in red is a regression line, with fitting parameters as shown, for the median values.

Both types of rainfall data were routinely ingested into MPI RSM through PI. To single out the contribution of radar rainfall data, satellite information was not assimilated in the 4DVAR Cycles. In the present impact study, MPI-RSM 4DVAR adopted a 3-hour assimilation window ending at analysis time with rainfall data ingested in 3 batches of hourly accumulated rainfall.

3.4. *QPF Verification*

As the observed and forecast rainfall data have different resolutions, pre-processing are needed before carrying out a grid-based forecast verification. To this end, all rainfall data are re-analyzed onto a 16 × 16 verification grid, each grid box being roughly 28 × 28 km^2. The total verification area is around 450 × 450 km^2, which is basically a square inscribed inside the 256-km range radarscope. Fig. 1 shows a map overlaid with the verification grid centred at Hong Kong.

Since the radar data are at a resolution much higher than 28 km, observed rainfall values within each verification grid box are up-scaled by taking an arithmetic mean. For forecast rainfall data, nearest model grid-point values are interpolated into a verification grid box. After such pre-processing, raw verification statistics including “hit”, “missed”, “false”, “correctly rejected” counts were computed from the three re-analyzed rainfall fields according to 5 chosen rainfall threshold values, namely 1, 5, 10, 20 and 30 mm/h. Threat Score and Bias Score were then calculated for examining forecast skill and bias respectively.

At first sight, the above chosen threshold values might appear to be not enough for representing heavy precipitation characteristics of southern China. However, it must be emphasized that over the scale of a verification grid box, sub-grid variability of observed rainfall was substantial. Fig. 2 shows the sub-grid radar rainfall distribution (middle 50 % range) against verification rainfall threshold. As illustrated in Fig. 2, the sub-grid median value increased linearly with verification threshold. The spread of sub-grid distribution also became significantly larger with increasing verification threshold. Considering the 5 mm/h threshold, Fig. 2 suggested that intensities of localized heavy downpours within a verification grid box could range from, say, 15 to 35 mm/h. As each grid box was roughly a quarter of the size of Hong Kong, it implied a finite chance to attain Green or even Amber rainstorm intensities (refer to Table I for definitions of rainfall intensity classes). For higher threshold values, intense rainstorms would be almost guaranteed. As a general deduction from Fig. 2, rainfall thresholds greater than or equal to 1 mm/h could imply heavy rainfall, at least over a local scale.

4. VERIFICATION RESULTS

For easy discussion, the statistical results were delineated into 3 forecast ranges, namely the very short range (VSR, 1-6 hours), intermediate range (IR, 7-18 hours) and short range (SR, 19-42 hours). Such stratification was contrived to allow for higher tolerance for timing errors as the forecast range lengthened.

4.1. *Rainfall Frequency Distribution*

The histograms of Fig. 3 show the frequency distributions of rain events under different threshold values. Observed radar rainfall, 4DVAR initialized precipitation forecasts and routine forecasts based on PI are shown in green, blue and red respectively. The frequency counts were compiled over the verification

grid as the number of grid boxes with rainfall amount greater than or equal to the threshold values. Effectively, such threshold values specified the minimal rainfall intensity over a verification grid box.

Apart from the VSR, there existed a general trend for forecast distributions to fall much too quickly than observations as rainfall threshold values increase (see histogram plots (b)-(d) of Fig. 3). This under-forecasting trend was also reflected in the Bias Score to be discussed in the Section 4.2. As a consequence of using radar information, forecast distributions have much higher counts relative to observations in the VSR. As shown in Fig. 3(a), the forecast distributions overshot the observed frequency at the 1 mm/h threshold but otherwise resembled closely to the observed trend during the first 6 hours. Such an overshooting might not necessarily imply poorer forecast skill as location was not taken into account in the frequency plot. In fact, QPF from 4DVAR Cycle scored higher probability of detection while maintaining similar false alarms comparing with Routine Cycle runs (results not shown).

The impact of 4DVAR might be inferred from its higher frequency counts recorded for VSR QPF relative to routine forecasts. This trend was clearly shown at all threshold values in this range. The ability of 4DVAR to produce more intense rainfall forecast is unambiguously seen from the significantly higher frequency counts for the 10 and 20 mm/h regimes.

The value of radar rainfall data might be inferred indirectly from the drastic drop in forecast frequency counts in both the IR and SR. With a limited spatial coverage (see Fig. 4(c) for the actual radar 256-km range domain) of the order of 500 $\times$ 500 km^2 only, radar data's influence is expected to vanish beyond, say, 10 hours under synoptic flow speeds typically found in May and June over southern China (say, about 50 km/h). Therefore, the VSR, IR and SR forecast ranges might be interpreted as three time windows in which radar data has strong, fading and vanishing influence respectively. The drop in frequency counts starting from the IR as shown in Fig. 3 (b) and (c) could then be attributed to the lack of reliable rainfall information outside the radar coverage.

In the IR, the lower frequency counts (Fig. 3(b) refers) compiled from 4DVAR-based forecasts for the 5 mm/h or higher thresholds are rather unexpected at the first glance. But as a frequency histogram only counts the number of grid boxes satisfying a given threshold value and does not take the spatial distribution of rainfall into consideration, lower frequency counts for particular threshold values only imply that the corresponding overall rain areas are smaller. Whether or not the forecast skills would become poorer depends critically on the detailed spatial distributions. Based upon subjective

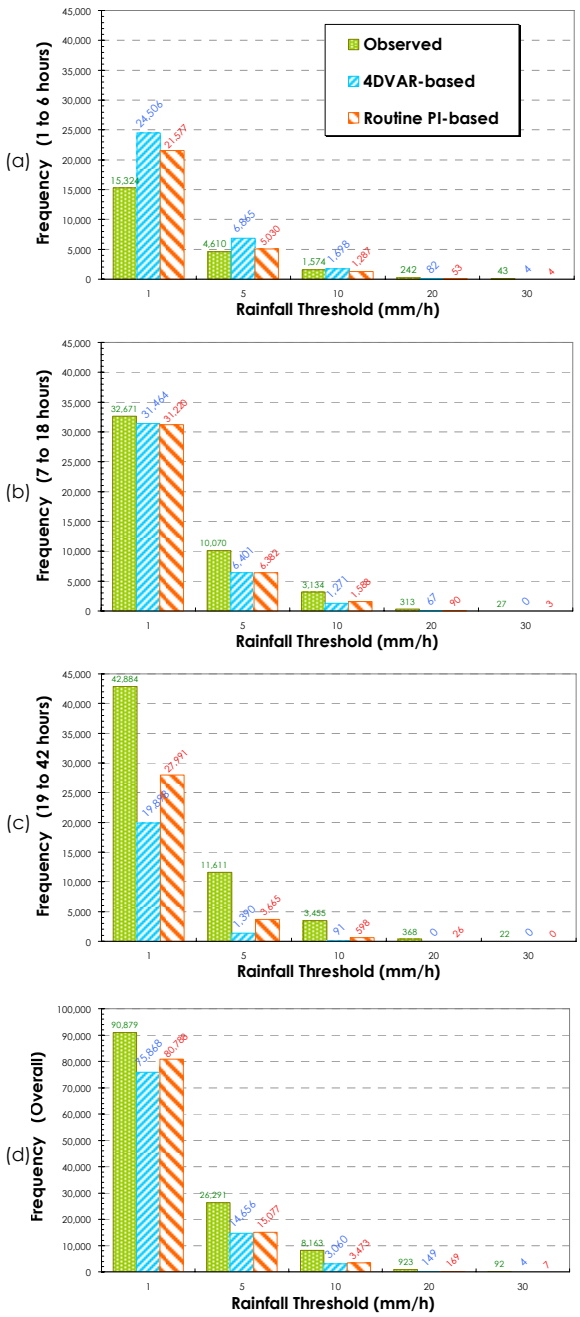
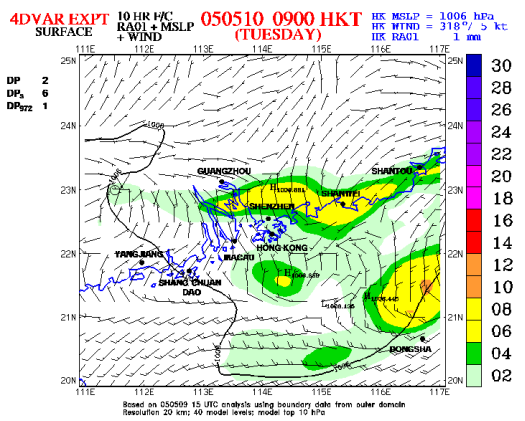


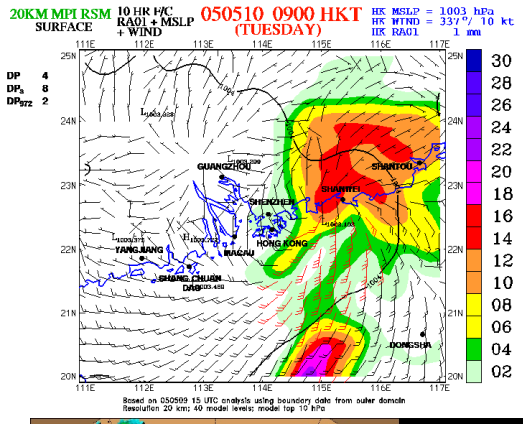
Fig. 3 Frequency distributions of rainfall events under threshold values 1, 5, 10, 20 and 30 mm/h, stratified into three forecast ranges: (a) VSR; (b) IR and (c) SR. Histogram (d) represents the overall frequency distributions irrespective of forecast range. In all histograms, frequencies for the observed, 4DVAR-based forecast and routine PI-based forecast are shown in green, blue and red respectively.

comparison on selected forecast rainfall fields, one possible explanation for the observed lower frequency counts for 4DVAR-based forecasts in the IR may be ascribed to the better mesoscale dynamical structures found in 4DVAR-based forecasts. The rainstorm case

(a)



(b)



(c)

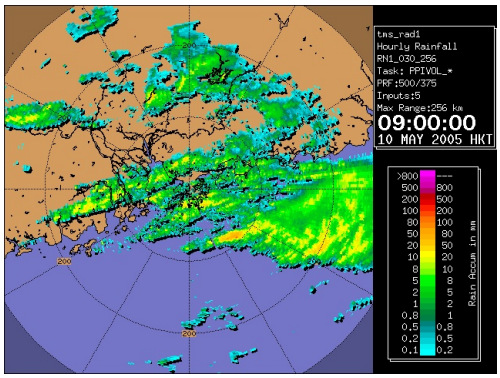


Fig. 4 Rainstorm on 10 May 2005: (a) better defined wind shear line and narrower rain bands resulted from 4DVAR analysis; (b) overly spread and intense rain bands due to erroneous dynamical structure given by routine PI-based forecast; (c) actual rain bands as seen by radar.

on 10 May 2005 may help illustrate this point. As depicted in Fig. 4(a), the 10-hour surface forecast based on 4DVAR gave a wind-shear line along the coast of Guangdong and resulted in a west-east oriented rain bands there. In contrast, the routine forecast based on PI gave an overly spread and intense rain pattern due to an erroneous mesocyclone development as shown in Fig. 4(b). Together with the re-analysis procedure in QPF verification, which tends to even out fine/narrow features in rain patterns, the

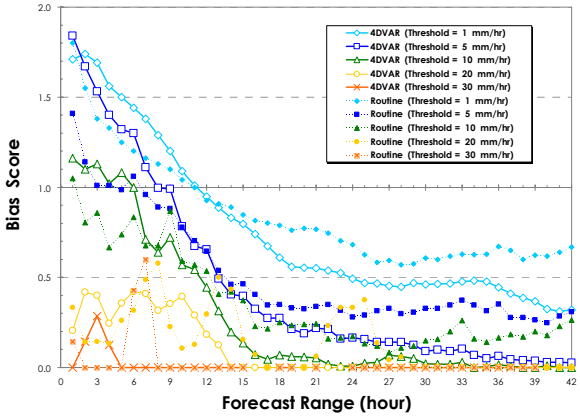


Fig. 5 Bias scores as functions of forecast range. Forecasts initialized by 4DVAR are shown as solid lines with open symbols whereas routine forecasts are marked by dashed lines with filled symbols. Five color scales ranging from light blue to orange, representing rainfall threshold values 1, 5, 10, 20 and 30 mm/h respectively.

rain areas and hence frequency counts for heavy rainfall could actually become smaller. The fact that the forecast skills (details to be given in Section 4.3) of 4DVAR-based forecasts were not significantly lowered at the 5-mm/h threshold in the first 18 hours was a reflection that 4DVAR actually improved and not degraded QPF at such rainfall intensity.

The significant difference in the SR frequency counts between 4DVAR-based and PI-based forecasts as shown in Fig. 3 (c) was primarily due to the use of satellite-estimated rainfall by PI and lack of rainfall information for 4DVAR outside the radarscope. Such difference in frequency counts did not necessarily imply deterioration in forecast skill as supported by the observed trends in forecast skills (details to be given in Section 4.3). Conversely, the ability to maintain compatible forecast skills even out to the SR in the absence of domain-wide rainfall information might be interpreted as a merit of 4DVAR analysis.

4.2. *Frequency Bias*

Fig. 5 plots the Bias Score (BS) as a function of forecast range. For better appreciation of the impact on frequency bias, we again stratified the statistics into the VSR, IR and SR forecast ranges and plotted the resulting BS as a function of rainfall threshold in Fig. 6. As evident in both Fig. 5 and Fig. 6 (b)-(d), both forecast data sets showed an overall trend of under-predicting the occurrence of rain events for all rain intensities beyond the VSR. Such bias was particularly serious in the more intense rain regime. As Fig. 5 and Fig. 6 (a)-(c) depicted, the under- prediction problem deteriorated as the forecasting range lengthened. For lighter rain intensity in VSR, there was an opposite bias trend of over-prediction.

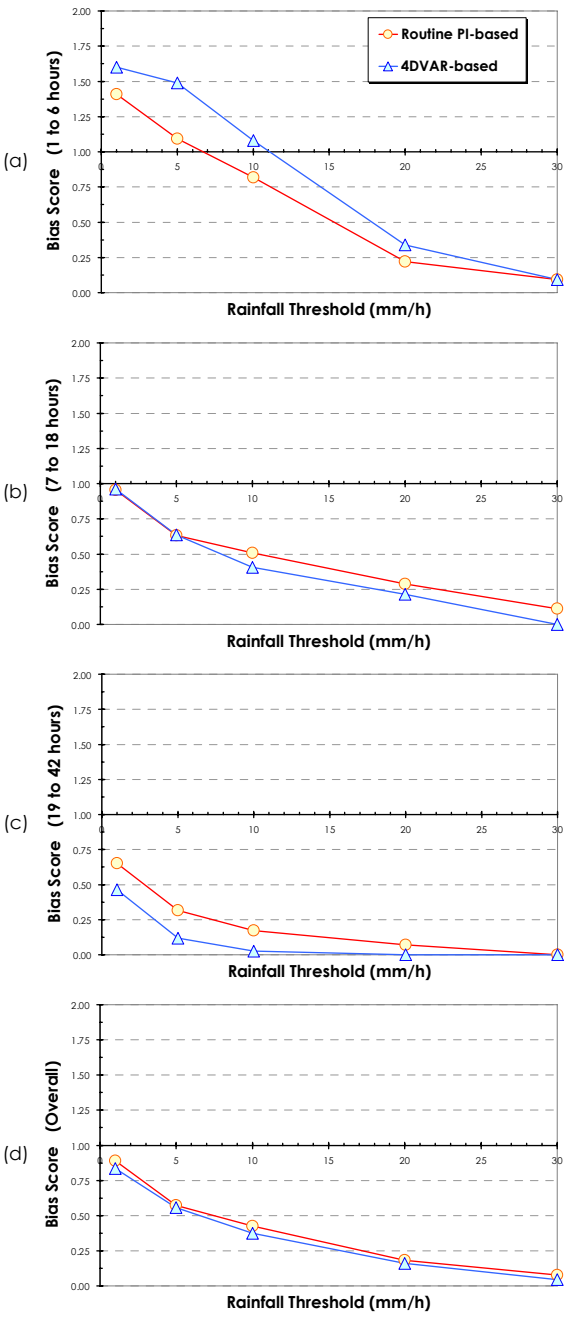


Fig. 6 Bias Score versus rainfall threshold, stratified into three forecast ranges: (a) VSR; (b) IR and (c) SR. Histogram (d) represents the overall frequency bias independent of forecast range. Forecasts initialized by 4DVAR and PI (routine) are shown in blue and red respectively.

As a result, the 4DVAR-based forecasts were relatively un-biased (i.e. BS ~ 1) in two regimes only: (i) 10 mm/h in the VSR and (ii) 1 mm/h in the IR. The routine PI-based forecasts were also unbiased for 5 mm/h in the VSR. For more intense rainfall, 4DVAR-based forecasts had relatively less bias in the VSR but behaved oppositely in the IR and SR.

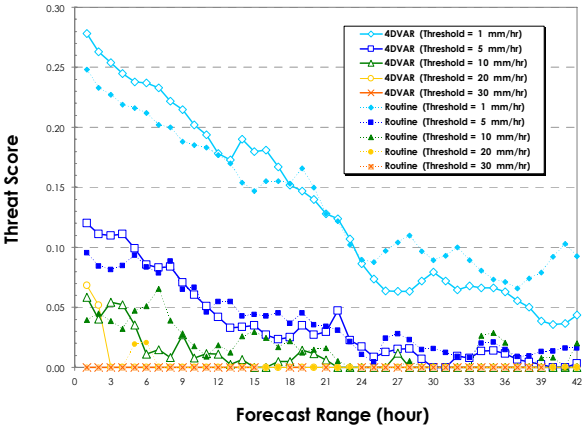


Fig. 7 Threat scores as functions of forecast range. Forecasts initialized by 4DVAR are shown as solid lines with open symbols whereas routine forecasts are marked by dashed lines with filled symbols. Five color scales ranging from light blue to orange, representing rainfall threshold values 1, 5, 10, 20 and 30 mm/h respectively.

Possible explanations were mentioned in Section 4.1.

4.3. *Forecast Skill*

Threat Score (TS) was chosen and presented below for the assessment of forecast skill. Other skill indicators, such as Equitable Threat Score (ETS) and Hanssen-Kuipers Score (HKS), have also been evaluated. Due to our choice of rain-oriented data sets, ETS and HKS did not show significant difference from TS and would not be discussed further in this paper. Fig. 7 plotted TS as a function of forecast range. In general, forecast skill deteriorated on increasing forecast range or rain intensity, becoming small at 20 mm/h or higher. As the frequency count for each individual forecast hour is not high and becomes even less significant statistically in the more intense rain regime, random fluctuations are seen in the plot of Fig. 7, especially for the 10 mm/h or higher threshold.

For better appreciation of the forecast skills, the raw statistics were stratified into the VSR, IR and SR forecast ranges and plotted against rainfall thresholds in Fig. 8(a), (b) and (c) respectively. Due to the use of radar data, the skills of both 4DVAR-based and PI-based forecasts peaked in the VSR. The advantage of 4DVAR analysis technique was prominent in this range as seen from the consistently higher TS for all rainfall threshold values (see Fig. 8(a)). For the more intense regime at 20 mm/h in the VSR, 4DVAR out-performed PI by nearly a factor of three.

Fig. 8 (b) and (c) showed that the impact of 4DVAR analysis on QPF skill extended to the IR at the 1 mm/h threshold but appeared not progressing any further beyond that regime. As explained in Section 4.1, forecasts from the 4DVAR Cycles, which by construction relied purely on radar for rainfall

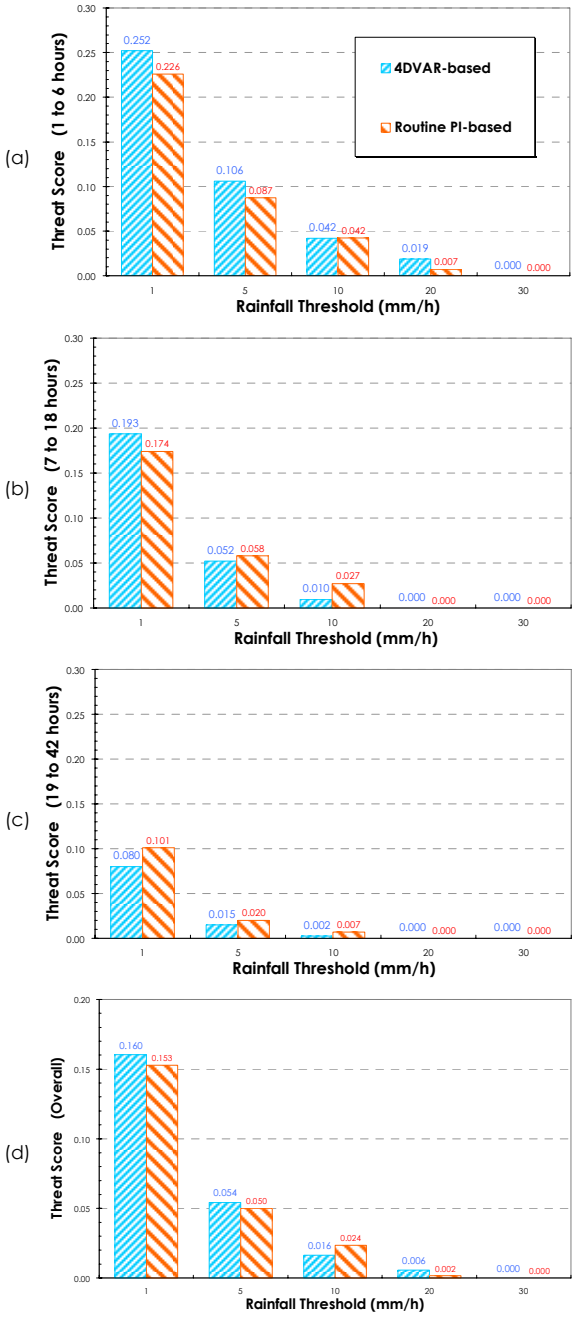


Fig. 8 Threat score in May and June 2005 versus rainfall threshold, stratified into three forecast ranges: (a) VSR; (b) IR and (c) SR. Histogram (d) represents the overall skill level independent of forecast range. Forecasts initialized by 4DVAR and PI (routine) are shown in blue and red respectively.

information, may not be more skillful in the IR and SR than those from Routine Cycle, which ingested both radar- and satellite-based rainfall information.

5. SUMMARY & CONCLUSION

Despite the range of forcing mechanisms and rain intensity among the chosen cases, the objective QPF verification results indicated that the 4DVAR analysis technique in general produced superior results in both the analysis and forecast precipitation fields, especially in the early hours of simulation.

Case studies of three selected heavy rain events (results not shown) also revealed that more realistic analysis from the 4DVAR technique produced forecast fields that were able to depict the underlying forcing mechanisms for the observed intense convection. The associated mesoscale features were better organized and aligned with respect to the radar reflectivity and satellite images. Convection in terms of peak intensity and the relative positioning of rain areas were also better represented and forecast using the 4DVAR technique.

In conclusion, the present study fully demonstrated the capability of 4DVAR technique in assimilating radar-based rainfall data and its significant positive impact on VSR QPF. To predict more successfully with even longer lead time, reliable rainfall information would be important. Further systematic numerical experiments based on satellite rainfall data assimilation and objective QPF verification would be required to assess if 4DVAR technique was equally useful out to the short range.

REFERENCE

JMA, 2002: Outline of the Operational Numerical Weather Prediction at the Japan Meteorological Agency. Appendix to WMO Numerical Weather Prediction Progress Report, 158.

Koizumi, K., *et al*, 2003: Improvement in precipitation forecasts of the JMA mesoscale model with four-dimensional variational data assimilation. *Proceedings of International Workshop on NWP Models for Heavy Precipitation in Asia and Pacific Areas*, 4-6 February 2003, Tokyo, Japan, 122-128.

Matsumura, T., *et al*, 1995: Improvement of spin-up of precipitation calculation with use of observed rainfall in the initialization scheme. *J. Met. Soc. Jpn.*, **73**, 353-368.

Tsuyuki, T., *et al*, 2002: The JMA Mesoscale 4D-Var System and Assimilation of Precipitation and Moisture Data. *ECMWF Workshop*, *8-11 July 2002*.

ACKNOWLEDGEMENT

The authors would like to thank Mr. K.H. Yeung for his valuable comments and suggestions.