Vietnam Journal of Earth Sciences, 40(4), 367-394, Doi: 10.15625/0866-7187/40/4/13134

(VAST)

Vietnam Academy of Science and Technology

Vietnam Journal of Earth Sciences

http://www.vjs.ac.vn/index.php/jse

Application of ensemble Kalman filter in WRF model to forecast rainfall on monsoon onset period in South Vietnam

Pham Thi Minh*, Bui Thi Tuyet, Tran Thi Thu Thao, Le Thi Thu Hang

Department of Meteorology, Hydrology and Climate change, Ho Chi Minh University of Natural Resources and Environment, 236B Le Van Si street, Ward 1, Tan Binh District, Ho Chi Minh City

Received 07 June 2018; Received in revised form 31 August 2018; Accepted 25 September 2018

ABSTRACT

This paper presents some results of rainfall forecast in the monsoon onset period in South Vietnam, with the use of ensemble Kalman filter to assimilate observation data into the initial field of the model. The study of rainfall forecasts are experimented at the time of Southern monsoon outbreaks for 3 years (2005, 2008 and 2009), corresponding to 18 cases. In each case, there are five trials, including satellite wind data assimilation, upper-air sounding data assimilation, mixed data (satellite wind+upper-air sounding data) assimilation and two controlled trials (one single predictive test and one multi-physical ensemble prediction), which is equivalent to 85 forecasts for one trial. Based on the statistical evaluation of 36 samples (18 meteorological stations and 18 trials), the results show that Kalman filter assimilates satellite wind data to forecast well rainfall at 48 hours and 72 hours ranges. With 24 hour forecasting period, upper-air sounding data assimilation and mixed data assimilation experiments predicted better rainfall than non-assimilation tests. The results of the assessment based on the phase prediction indicators also show that the ensemble Kalman filter assimilating satellite wind data and mixed data sets improve the rain forecasting capability of the model at 48 hours and 72 hour ranges, while the upper-air sounding data assimilation test produces satisfactory results at the 72 hour forecast range, and the multi-physical ensemble test predicted good rainfall at 24 hour and 48 hour forecasts. The results of this research initially lead to a new research approach, Kalman Filter Application that assimilates the existing observation data into input data of the model that can improve the quality of rainfall forecast in Southern Vietnam and overall country in general.

Keywords: WRF-LETKF; Nambo; Rainfall; Heavy rainfall; Data assimilation; Forecast.

1. Introduction

Coastal zone has a hugely significant role in Vietnam socio-economic development. However, the region is facing with natural hazards like coastal erosions, typhoons and climate change.

Nowadays, rainfall forecast by numerical model is still a challenging problem, because rainfall is an unstable and intermittent meteorological element. There are many research applications of numerical models in the world for rain forecast and forecast results have improved significantly. Recently, some authors investigated the possibility of heavy rain forecasts of 72 hour range or more (Schumacher and Davis, 2010, Lin et al., 2010), but the accuracy of the rain intensity forecasts of these studies is unreliable. However, other authors

*Corresponding author, Email: minhpt201@gmail.com

used models which well predicted rainfall in high-resolution (1.5 km) at 24 h and 3 h ranges (Elemeti et al., 2005; Kato and Aranami, 2009). In addition, a number of other studies focused primarily on improving the initial field of the model by assimilation methods to predict rainfall. For example, Xavier (2006) used the 3DVAR assimilation method to assimilate the temperature and humidity profiles from MODIS satellites and radio upper-air sounding data in the initial field of the MM5 model. Xavier's research has shown that the use of MODIS satellite data has significantly improved the occurrence of heavy rainfall associated with tropical depression. And Routray et al. (2008) also used the 3DVAR method to assimilate surface weather stations, ship, buoy, upper-air sounding and Kapanal-1 geostationary satellite data to forecast heavy rainfall in India. Routray's calculations show that the data assimilation has significantly improved the quality of heavy rainfall simulations in India at 24 hour and 48 hour ranges.

In Vietnam, some rainfall predicting studies by numerical methods also achieved important results. Specifically, Tran Tan Tien and Nguyen Thi Thanh (2011) forecast heavy rainfall in the Central region using the WRF model with the 3DVAR approach assimilating MODIS satellite data into the input field of the model. This study has predicted the rainfall region and rainfall measurement at the station in accordance with observation. In particular, this study simulates accumulated precipitation of 400 mm/72 hours. Other authors have used the HRM, MM5 and WRF models for large area heavy rainfall forecasts in Vietnam with positive results (Kieu Thi Xin, 2005, Hoang Duc Cuong, 2008, Kieu Thi Thuy et al., 2013). Most rainfall studies in Vietnam are concentrated in central Vietnam (Nguyen Duc Phuong, 2013, Cong Thanh et al., 2015, Nguyen Khanh Van et al., 2013; Simon Wang and et al., 2014).

However, the physical processes in the modern models are quite perfect. Therefore, the forecast error cannot be reduced. Thus, to reduce the predicted error, some recent studies focus primarily on improving the input field of the regional model by data assimilation methods. Assimilation of the data goes in two directions: variational and sequencing assimilations. In particular, the sequence assimilation method is strongly developed, because this method is easy programming and easy application for any nonlinear model. Representation of sequence assimilation method is ensemble Kalman filter (Kalnay, 2003). And one variant of the ensemble Kalman filter is the Local Ensemble Transform Kalman Filter (LETKF) which is widely used in global prediction models (Hunt et al., 2007; Miyoshi and Yamane, Szunyogh et al., 2008) and regional models (Miyoshi and Kunii, 2012; Kieu et al., 2012; Maria E. Dillon, 2015). These studies indicate that LETKF generates a more accurate analysis field than the analysis field of other methods.

In Vietnam, the sequence assimilation method is still a new issue. Kieu Quoc Chanh (2011) investigated and applied the Local Ensemble Transform Kalman Filter (LETKF) in the WRF model. The results of this study suggest that the Local Ensemble Transform Kalman Filter is capable of capturing well the satellite wind data. In addition, Nguyen Duc Phuong (2013) used LETKF which assimilates upper-air sounding data to heavy rainfall forecasts in Central Vietnam. The results show that LETKF has certain advantages in heavy rainfall forecast. Thus, in this study we used the Local Ensemble Transform Kalman Filter applied in the WRF model to predict rainfall in the South of Vietnam.

On the other hand, the monsoon onset is a period of complex conflict between large-scale systems (Bui Minh Tuan and Nguyen Minh Truong, 2013, Nguyen Minh Truong and Bui Minh Tuan, 2013). This fighting changes the meteorological elements. Therefore, weather forecast and rain forecast are

very difficult. Previous studies have focused only on the monsoon outbreak (Tanaka, 1992; Wang and Wu, 1997; Wang, 2004; Fasullo and Webster, 2003) and the monsoon onset area (Tao and Chen, 1987; Wu and Zhang, 1998; Li C. and Qu X., 1999; Zhang and ccs, 2004; Lau, K. M., and S. Yang, 1997; Matsumoto J., 1997; Webster et al., 1998; Wang B., and Z. Fan, 1999; Lu J., Zhang Q., Tao S., and Ju J., 2006; He J., Yu J., Shen X., and Gao H., 2004; Wang B., 2003). Forecasting rain in this period has not been much attention. In addition, rainfall is one of the factors used to determine the timing of a monsoon onset and it is important to forecast accurate rainfall during the monsoon onset period. Therefore, we chose the rain forecast in the monsoon onset period to assess the predictability of the WRF model when applied Local Ensemble Transform Kalman Filter to assimilate observation data.

2. Ensemble Kalman filter and estimate methods

2.1. Ensemble Kalman filter

Generally, data assimilation is the best initial field-generating process possible for a predictive model, based on motivational relationships and statistical probabilities. Due to the specificity of the weather models using numerical models that strongly depends on the initial conditions, weather forecasts sometimes has the incorrect results due to inaccurate initial conditions (LF Richardson, 1922; GJ Halriner and RT Williams, 1982; TN Krishnamurti, L. Bounoa, 1996; E. Kalnay, 2003; N. Philips, 1960).

The data assimilation approach includes 2-steps: (1) Objective analysis and (2) Initialization (R Daley, 1991). According to the objective analysis, the observation field will be optimally extrapolated to the grid point of the model. In the initialization, this extrapolation field will be balanced so that the interdependent observation variables will be bound by a given dynamic relation, this is necessary to avoid arbitrary observation values. The data assimilation has been around since the 1950s (J.G. Charney, 1955; N.A. Phillips, 1960b), but for the last 20 years, the data assimilation has actually grown. From a modern point of view, the current data assimilation schemes can be divided into two categories: variational and sequential approaches (P. Courtier, O. Talagrand, 1987). That sequential approach has certain advantages in programming and application convenience in the models. Typical sequential assimilation approach is Kalman filter assimilations (C. Snyder, F. Zhang, 2003; M.K. Tippett, 2003; J.S. Whitaker, T.M. Hamill, 2002; TMJS Whitaker, C. Snyder, Houtekamer, 2005; B.R. Hunt, E. Kostelich, I. Shunyogh, 2007); given demonstrated capability of a variant of the Ensemble Kalman filter algorithm, the so-called Local Ensemble Transform Kalman Filter (LETKF) (see, e.g., Miyoshi and Kunii, 2012; Kieu et al., 2012; Maria E. Dillon, 2015).

The LETKF algorithm use a background ensemble matrix as a transformation operator from the model space spanned by the grid points within a local patch to an ensemble-space spanned by the ensemble numbers, and to perform an analysis in this ensemble space at each grid points. For the LETKF algorithm, it assumed that there is a ensemble background {xb (i): i = 1,2..., k}, where k is the number of ensembles. According to Hunt et al. (2007), an ensemble averaging matrix and a noise matrix are defined:

$$\overline{\boldsymbol{x}}^b = \frac{1}{k}\sum_{i=1}^{k} \boldsymbol{x}^{b(i)}$$
$$\boldsymbol{X}^b = \boldsymbol{x}^{b(i)} - \overline{\boldsymbol{x}}^b \quad (1)$$

Symbol $\boldsymbol{x} = \overline{\boldsymbol{x}}^b + \boldsymbol{X}^b\boldsymbol{w}$, where $\boldsymbol{w}$ is a local vector, the localized cost-function is minimized in the form:

$$\widehat{J}(\mathbf{w}) = (k-1)\mathbf{w}^T\{\mathbf{I} - (\mathbf{X}^b)^T[\mathbf{X}^b(\mathbf{X}^b)^T]^{-1}\mathbf{X}^b\}\mathbf{w} + J[\mathbf{x}^b + \mathbf{X}^b\mathbf{w}] \quad (2)$$

In this case $J[\mathbf{x}^b + \mathbf{X}^b\mathbf{w}]$ is the cost function in the model space. If the cost function determines in the full space of $\mathbf{X}^b$ ($N = \{\mathbf{v} | \mathbf{X}^b\mathbf{v} = 0\}$), it is easy to see that is divided into two parts: This includes the composition of $\boldsymbol{w}$ in N (the first term in equation 2), and the second component depends on the composition of $\boldsymbol{w}$ orthogonal to N. The analytical state conditions are orthogonal to N to the minimized, the average analysis state matrix, and its corresponding correlation matrix in the composite space can be expressed as:

$$\overline{\mathbf{w}}^a = \widehat{\mathbf{P}}^a(\mathbf{Y}^b)^T\mathbf{R}^{-1}[\mathbf{y}^0 - H(\overline{\mathbf{x}}^b)] \quad (3)$$

$$\widehat{\mathbf{P}}^a = [(k-1)\mathbf{I} + (\mathbf{Y}^b)^T\mathbf{R}^{-1}\mathbf{Y}^b]^{-1} \quad (4)$$

Where $\mathbf{Y}^b \equiv H(\mathbf{x}^{b(i)} - \overline{\mathbf{x}}^b)$ is matrix of observation errors background, R is observations error covariance matrix. Note that, analysis error covariance matrix $\mathbf{P}^a$ in the space model and analysis error covariance in the ensemble space $\widehat{\mathbf{P}}^a$ is relationship: $\mathbf{P}^a = \mathbf{X}^b\widehat{\mathbf{P}}^a(\mathbf{X}^b)^T$, analysis ensemble pertubation matrix $\mathbf{X}^a$ can be expressed as:

$$\mathbf{X}^a = \mathbf{X}^b[(k-1)\widehat{\mathbf{P}}^a]^{1/2} \quad (5)$$

Analysis ensemble $\mathbf{x}^a$ can be calculated as following:

$$\mathbf{x}^{a(i)} = \overline{\mathbf{x}}^b + \mathbf{X}^b\{\overline{\mathbf{w}}^a + [(k-1)\widehat{\mathbf{P}}^a]^{1/2}\} \quad (6)$$

More details on the LETKF algorithm can be found in Hunt et al. (2007).

Where, $\overline{\mathbf{x}}^b$ is averaged ensemble matrix; $\mathbf{X}^b$ The ensemble pertubation matrix; $\widehat{\mathbf{P}}^a$ analysis error covariance matrix and $\overline{\mathbf{w}}^a$: weight matrix in ensemble space.

2.2. Estimate approach

In this study, we used statistical evaluation indicators such as mean error (ME), mean absolute error (MAE), and root mean square error (RMSE) (Wilks, 1997).

2.2.1. Statistical evaluation indicators

Mean error ME: $ME = \frac{1}{n}\sum_{i=1}^{n}(F_i - O_i)$ (7)

Where: n is number of forecast; F_i is the i forecast; Oi is the i observation value.

Mean absolute error MAE:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|F_i - O_i| \quad (8)$$

Where: n is number of forecast, F_i is the i forecast; Oi is the i observation value.

Root mean square error RMSE:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Fi - Oi)^2} \quad (9)$$

Where: n is number of forecast, F_i is the i forecast; O_i is the i observation value.

RMSE is used to present the average magnitude of the error. Like MAE, RMSE does not reflect the deviation trend between the forecast value and the observed value. The optimal RMSE value is RMSE = 0, ie when the forecast value is equal to the observed value at every point in the evaluation space (Wilks, 1997).

2.2.2. Indicators for predicting phenomena/ phases

Table 1 illustrates how the observed frequencies are calculated for binary predictive variables (or two phases) where A is the total number of successful predictions (B is the predicted probability of occurrence). Similarly, on B occasions, the event was forecasted to occur but did not. There are also C instances of the event of interest occurring despite not being forecast, called misses; and D instances of the event not occurring after a forecast that it would not occur, sometimes called a correct negative. Based on the values of A, B, C and D in Table 1 we can calculate the indicators to evaluate the quality of the forecast. And Table 2 presents formulas, the

optimal values and value ranges of the four BIAS, POD, FAR and ETS indicators are used to evaluate the predictive power of the model.

The purpose of this study is the ensemble Kalman filter experiment to forecast the precipitation in monsoon onset period is the rainy events. According to the Circular "Technical regulation on hydro-meteorological forecasting procedures in normal conditions" of the Ministry of Natural Resources and Environment stipulates the time when rain is the days with the rainfall measurement more than 6 mm (R>3mm/12h).

Table 1. The Relationship between counts (letters A-D) of forecast/event pairs for the dichotomous nonprobabilistic verification situation

		Observed	
		Yes	No
Forecast	Yes	A	B
	No	C	D

Table 2. List of skill scores

Index	Calculation	Value arange	The best	Meaning
BIAS	$BIAS = \frac{A+B}{A+C}$	0 to $+\infty$	0	The bias is simply the ratio of the number of yes forecasts to the number of yes observation.
POD	$POD = \frac{A}{A+C}$	0 to 1	1	POD is the ratio of false alarms to the total number of nonoccurrences of the event or the conditional relative frequency of a wrong forecast given that the event does not occur.
FAR	$FAR = \frac{B}{A+B}$	0 to 1	0	FAR is the fraction of right forecasts that turn out to be wrong or proportion of the forecast events that fail to materialize.
ETS	$ETS = \frac{A - A_{random}}{A+B+C-A_{random}}$ $A_{random} = \frac{(A+C)(A+B)}{N}$	-1/3 to 1	1	Equitable threat score which is the ratio of success.

3. Description of experiments

3.1. Model

The purpose of the study was to evaluate the predictability of rain of data assimilation experiments. And the data used to assimilate in the input field of the model to be upper-air sounding data in the Asia and Vietnam, and satellite wind data. So the domain is defined as follows: the center domain is set at 12°N - 106.5°E; net 1 covers the area from 13°S-37°N and 81.5°E-131.5°E; Zone 2 covers the area from 4.1°N to 18.7°N and from 98.5°E to 113.5°E. The highest pressure level (upper margin of the model) is 10 hPa corresponding to the 31 level of the model. Horizontal resolution is 54/18 km (see Figure 1), the forecast range is 72h. Boundary conditions are updated every 6 hours from the GFS global forecasting model.

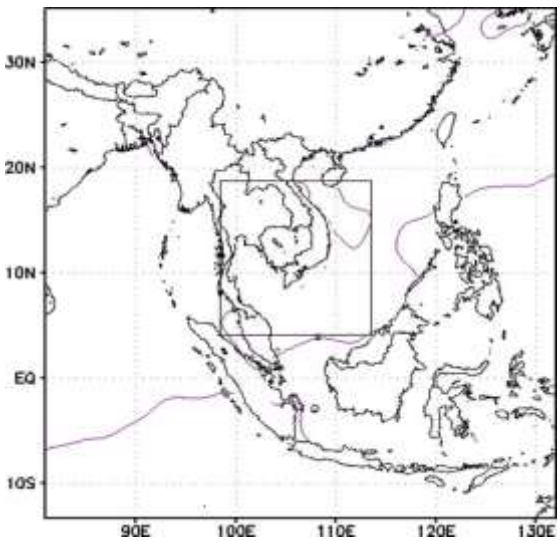

Figure 1. Forecasting domain Structure

In this study, the WRF model version 3.1.1 is applied, corresponding to micro-physical parameterization schemes, parameterization of longwave and shortwave radiation, convective parameterization, and their options are listed in Table 3. Control prediction with initial conditions and boundary conditions are taken from the GFS global model (CTL) and physical parameterization schemes include: WSM 3-class simple ice scheme, long wave rating scheme, YSU boundary layer parameterization scheme and BetetS Miller Janjic scheme; The multi-physical ensemble control (MPH) prediction and assimilation ensemble predictions with 21 components for physical parameterization schemes are given in Table 4.

Table 3. List of physical parameterization schemes in WRF model

Physical parameterization scheme	variables	Option
Microphysics	mp_physics	= 1, Kessler scheme
		= 2, Lin et al. scheme
		= 3, WSM 3-class simple ice scheme
		= 4, WSM 5-class scheme
		= 5, Ferrier (new Eta) microphysics
		= 6, WSM 6-class graupel scheme
Shortwave radiation	ra_sw_physics	= 1, Dudhia scheme
		= 2, Goddard short wave
Longwave radiation	Ra_lw_ physics	=1, rrtm scheme
Convection	cu_physics	= 1, Kain-FriETSch (new Eta) scheme
		= 2, BetETS-Miller-Janjic scheme

Table 4. The ensemble members and corresponds to the different physical parameterization scheme options

Members	Ra_lw_ physics	Ra_sw_ physics	mp_ physics	cu_physics
001	1	2	1	1
002	1	1	1	2
003	1	2	1	2
004	1	1	2	1
005	1	2	2	1
006	1	1	2	2
007	1	2	2	2
008	1	1	3	1
009	1	2	3	1
010	1	1	3	2
011	1	2	3	2
012	1	1	4	1
013	1	2	4	1
014	1	1	4	2
015	1	2	4	2
016	1	1	5	1
017	1	2	5	1
018	1	1	5	2
019	1	2	5	2
020	1	1	6	1
021	1	2	6	1

In addition, the author uses the LETKF scheme which applied in the research and Forecast Weather model (WRF) which is called WRF-LETKF system, was developed by Kieu (2011, HMO Faculty, VNU). The diagram of the WRF-LETKF system is illustrated in Figure 2.

Whereas the Advanced Research WRF (ARW) modeling is the most popular weather forecasting model today; WPS is a WRF pre-

processing system that allows interpolation of global prediction data in the resolution of the regional model; WRFDA is WRF Data Assimilation. The WRF model and related components are introduced in detail at http://www2.mmm.ucar.edu/wrf/users/docs/user_guide_v4/contents.html; LETKF (Local Ensemble Transform Kalman Filter) is a version's ensemble Kalman filter, whose algorithm introduced in section 2.1.

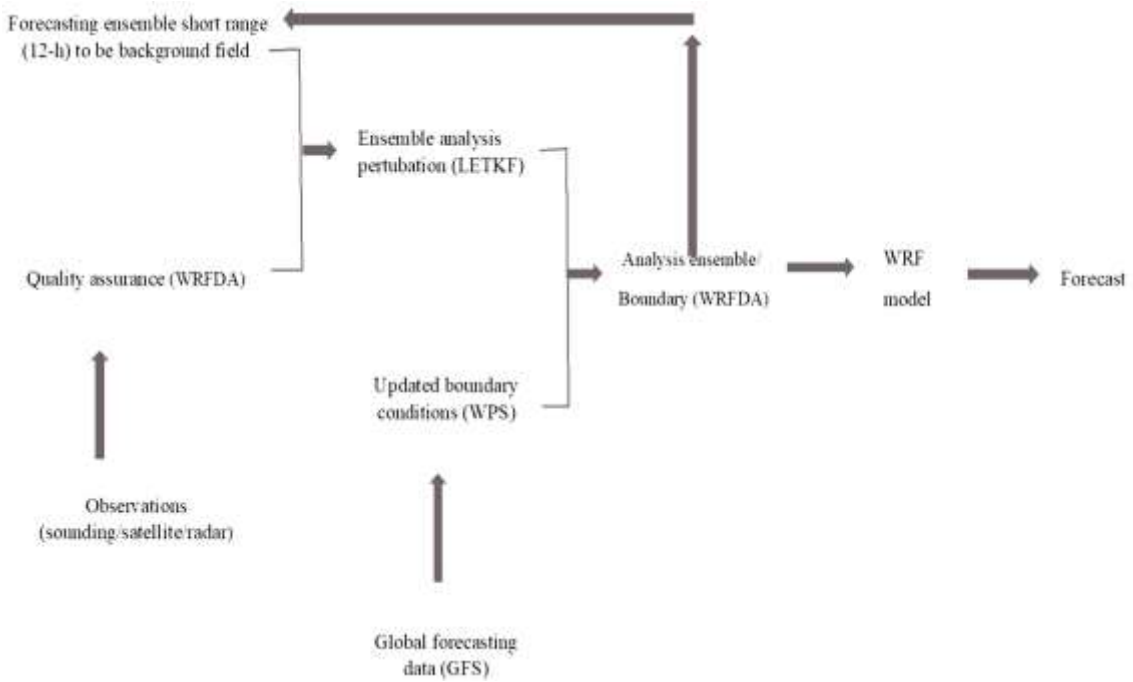

Figure 2. Diagram illustrating WRF-LETKF Ensemble Prediction System (Source: Kieu, 2011)

3.2. Numerical experiments

In this study, the author experimented with three cases which occurred near the time of the Southern monsoon outbreak in 2005, 2008 and 2009 years. These years were the rainy time that is near monsoon onset periods which lasted from 3 to 7 days. The run period is listed in Table 5.

Table 5. Experiment periods and starting forecast times

Year	Starting 12 h forecast	Starting 72 h forecast	Run numbers
2005	12 UTC 07 May	00 UTC 08 May	02
2008	00 UTC 26 April	12 UTC 26 April	11
2009	00 UTC 28 April	12 UTC 28 April	05

To evaluate the rainfall predictability of LETKF applied to the WRF model, we set up the experiments presented in Table 6. In which rain predictive experiments with the WRF model don't assimilate data (CTL) using boundary conditions and initial conditions is updated every 6 hours from the GFS global forecast model. The physical parameter options in data assimilated experiments and multi-physical ensemble experiment, see Table 3.

Table 4. List of experiments

Notation	Describing
CTL	Forecasting rainfall by non-assimilation WRF model
CIMSS	Forecasting rainfall by WRF model which assimilated satellite wind data with 21 ensembles
RADS	Forecasting rainfall by WRF model which assimilated upper-air sounding data with 21 ensembles
MIX	Forecasting rainfall by WRF model which assimilated mixed data (satellite wind data + upper-air sounding data) with 21 ensembles
MPH	Forecasting rainfall by WRF model which is 21 multi-physical ensembles

The rainfall prediction experiment with the WRF model assimilates observation data (satellite wind, upper-air sounding data, and mixed data) basically as follows: The first observed data will be processed Quality assurance through the WRFDA standard package in the WRF model. This quality assurance process will determine the errors for the levels and the corresponding observed variables. After the observed data will be combined with the 12 hour short-term forecast data from the previous prediction cycle to generate a set of analysis perturbation through the LETKF system. In this cycle, GFS global forecasting data is reported and downloaded to be preprocessed and interrupted in the region model grid. The GFS field will then be added to the analysis noise generated by the LETKF filter to generate an ensemble of the analytical fields along with the corresponding boundary conditions of the analyzed fields. The set of inputs and boundary generated in this step will be inserted into the WRF model to predict precipitation with a 72 hour forecast range. Parallel to this predetermined precipitation prediction, the WRF model will also store a set of forecasts as the background field for continue forecast. The ensemble forecasting process is repeated twice a day at 00 UTC and 12 UTC and after 12 hours the analysis cycle is performed. The start of the 72 hour rainfall forecast and the start of the short-term forecast period (12 hours) for the first analytical cycle are given in Table 5.

3.3. Data

3.3.1. Global Forecast System (GFS) data and upper-air sounding data

GFS Global forecasts data in the grib2 format with a resolution of 1° × 1° is used as boundary conditions and initial conditions in the experiments. Where the boundary conditions are updated every 6 hours. And upper-air sounding data is taken from stations in Asia and Vietnam regions (http://weather.uwyo.edu/upperair/sounding.html). For upper-air sounding data, the moist, heat and wind factors are entered to assimilate on the base levels (1000, 850, 700, 500, 400, 300, 200, 150, 100 and 50 hPa) and some minor levels.

Upper-air sounding data which collected for the experiments in this study are listed in Table 7, including the number of data collection stations and the number of stations with qualifying data for assimilation. Due to the observation data have to examine qualitative through the OBSPROC program of the WRFDA assimilation system. For example, at 00 UTC 8 May 2005, the total number of stations collected was 139, which is shown in Figure 3, after passing the OBSPROC quality control program, the remaining number of stations was 67. The distribution of 67 stations is shown in Figure 4.

Table 7. Numbers of upper-air sounding station for 18 experiments

Numbers	Time of forecast	Initial station numbers	Assimilated station numbers
1	00 UTC 2005/5/08	139	67
2	12 UTC 2005/5/08	123	53
3	12h-26/4/2008	110	65
4	00h-27/4/2008	83	67
5	12h-27/4/2008	96	67
6	00h-28/4/2008	100	61
7	12h-28/4/2008	105	61
8	00h-29/4/2008	110	67
9	12h-29/4/2008	103	59
10	00h-30/4/2008	105	60
11	12h-30/4/2008	101	59
12	00h-01/5/2008	102	63
13	12h-01/5/2008	94	53
14	12h-28/4/2009	96	62
15	00h-29/4/2009	76	43
16	12h-29/4/2009	96	56
17	00h-30/4/2009	103	68
18	12h-30/4/2009	91	54

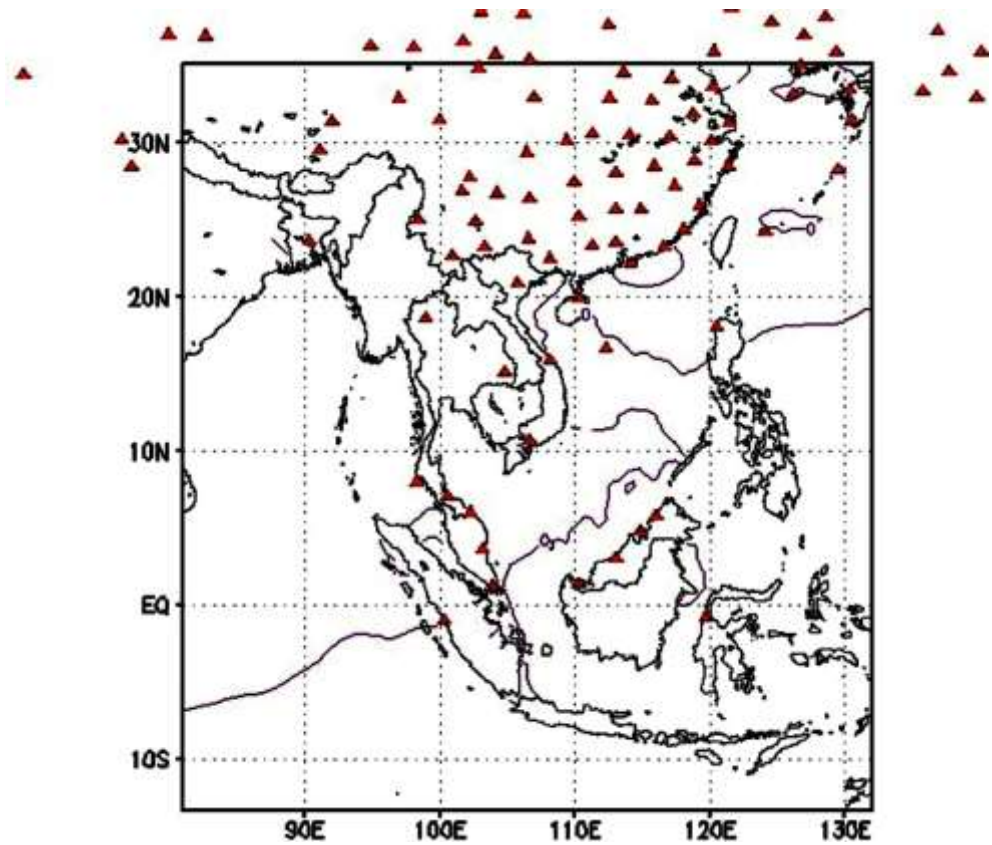

Figure 3. Distribution 139 upper-air sounding observed stations collected in Asian and Vietnam region

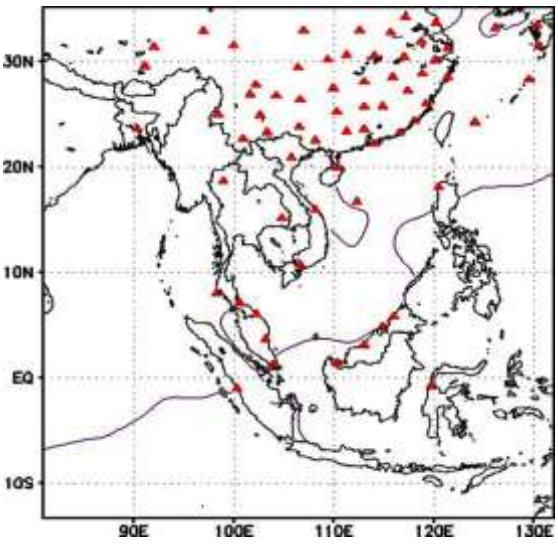

Figure 4. Distribution 67 eligible upper-air sounding observed stations which entered assimilated system at 00 UTC on 08 May 2005

3.3.2. The satellite wind data

The satellite wind data is a particularly important data source for operative forecasting models on the world. It covers the global and collected from 3 to 6 hours. It depends on the characteristics of each satellite. Satellite wind data describes the atmospheric dynamics structure that contributes to supplement the initial field data for the forecasting model by assimilation methods. Currently, satellite wind data is preprocessed by the Cooperative Institute for Meteorological Satellite Studies, University of Wisconsin satellite atmospheric motion vector (CIMSS-AMV) in the same time period selected. A number of studies with CIMSS-AMV data have shown that this data can help to improve the predictive quality of different medium-scale systems (Kieu et al., 2012, Miyoshi T. and Kunii M.). The advantage of CIMSS-AMV data is its error that is highly qualified and is determined by a recursive filter algorithm. Each data point is checked to match most of the surrounding data using the quality index technique. Most CIMSS-AMV data are distributed in different regions and are currently stored in a variety of formats including ASCII and/or BUFR. In this study, satellite wind data was collected in India, Pacific Northwest and Australia regions (Figure 5) and downloaded from http://tropic.ssec.wisc.edu.

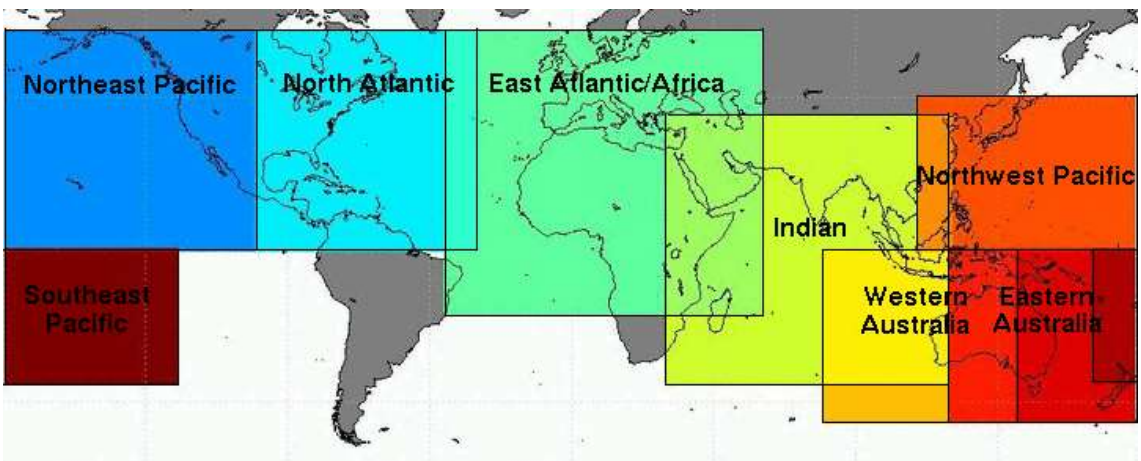

Figure 5. Regions where are covered by wind satellite (Source: http://tropic.ssec.wisc.edu)

Although satellite wind data is the input to the Global Data Assimilation System (GDAS), which produces the final tropospheric analysis, this analysis data is included in the global prediction model with rough resolution (1 degree) and are generally less predictable than observations. Thus, when using the product of the global prediction model as an input of the regional model, the interpolation process in the regional model disappeared large-scale simula-

tion information lead to the forecast results is not accurate, especially with rain forecast.

3.3.3. Observed rainy data at station

Observed rainfall data at the station are the important data taken from data sources of the National center Hydro-meteorological Forecasting. In fact, the Nam Bo area has 25 observation stations, but only 18 stations are used in this study (Table 8). With the permissible error of measuring the rainfall at the station is 0.1mm and is observed by modern equipment such as self-burners (SL1, SL3), self-recording type Symphony or by the staff conducts measured directly rainfall when observed by rain gauge. Rainfall at the stations was taken during the monsoon onset period of 2005, 2008, and 2009 years and is used as a benchmark to compare the results from the rain forecasting experiments.

Table 8. List of 18 synop stations at the Southern regions

The numbers	Station number	Station name	Longitude	Latitude
1	48883	Phuoc Long	107	11.8
2	48898	Tay Ninh	106.1	11.3
3	48900	Tan Son Hoa	106.7	10.8
4	48902	Ba Tri	106.6	10.1
5	48903	Vung Tau	107.1	10.4
6	48904	Cang Long	106.2	9.98
7	48906	Moc Hoa	105.9	10.8
8	48907	Rach Gia	105.1	10
9	48908	Cao Lanh	105.6	10.5
10	48909	Chau Doc	105.1	10.7
11	48910	Can Tho	105.8	10
12	48912	My Tho	106.4	10.4
13	48913	Soc Trang	106	9.6
14	48914	Ca Mau	105.2	9.18
15	48915	Bac Lieu	105.7	9.27
16	48917	Phu Quoc	104	10.2
17	48916	Tho Chu	103.1	9.27
18	48918	Con Dao	106.6	8.68

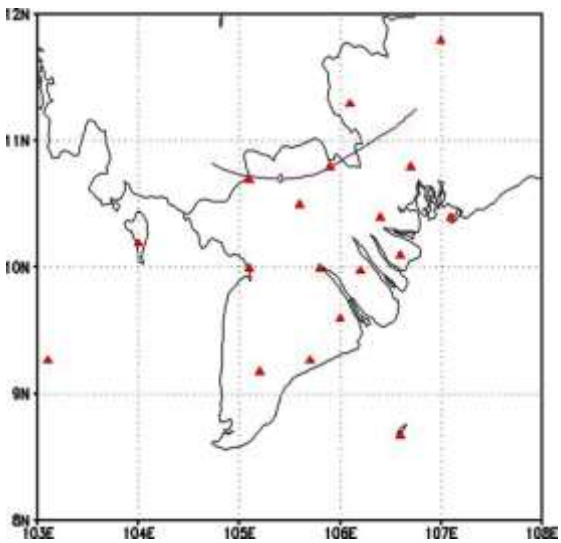

Figure 6. 18 Meteorology observed stations at the south of Vietnam

4. Results and Discussions

According to Nguyen Minh Truong and Bui Minh Tuan (2013), one of the most important mechanisms for large-scale convection outflows is wet convergence which is directly related to the wind field. Thus, wind field simulation results are one of the factors influencing rain simulation. Therefore, before assessing the predictability of rainfall of multi-physical ensemble Kalman filter, we investigate the effect of observed data on wind field simulation.

4.1. Influence of observation data on wind field simulation

For the wind field, we investigate the effect of observed data on the vertical structure of the wind components in rainfall forecast

experiments at 00 UTC 8 May 2005. In this case, the background field is derived from a 12 hour forecast with the starting time of forecast at 12 UTC 7 May 2005. The assimilation cycle begins at 00 UTC 08 May 2005, with the wind analysis increments (analysis field minus background field) and observed wind vectors shown in Figure 7. Figures 8 and 9 illustrate the initial wind field (background field), observed wind (black - satellite wind; red-upper-air sounding wind); And the analysis wind field which is made by LETKF assimilated satellite wind data and upper-air sounding data (green wind vector).

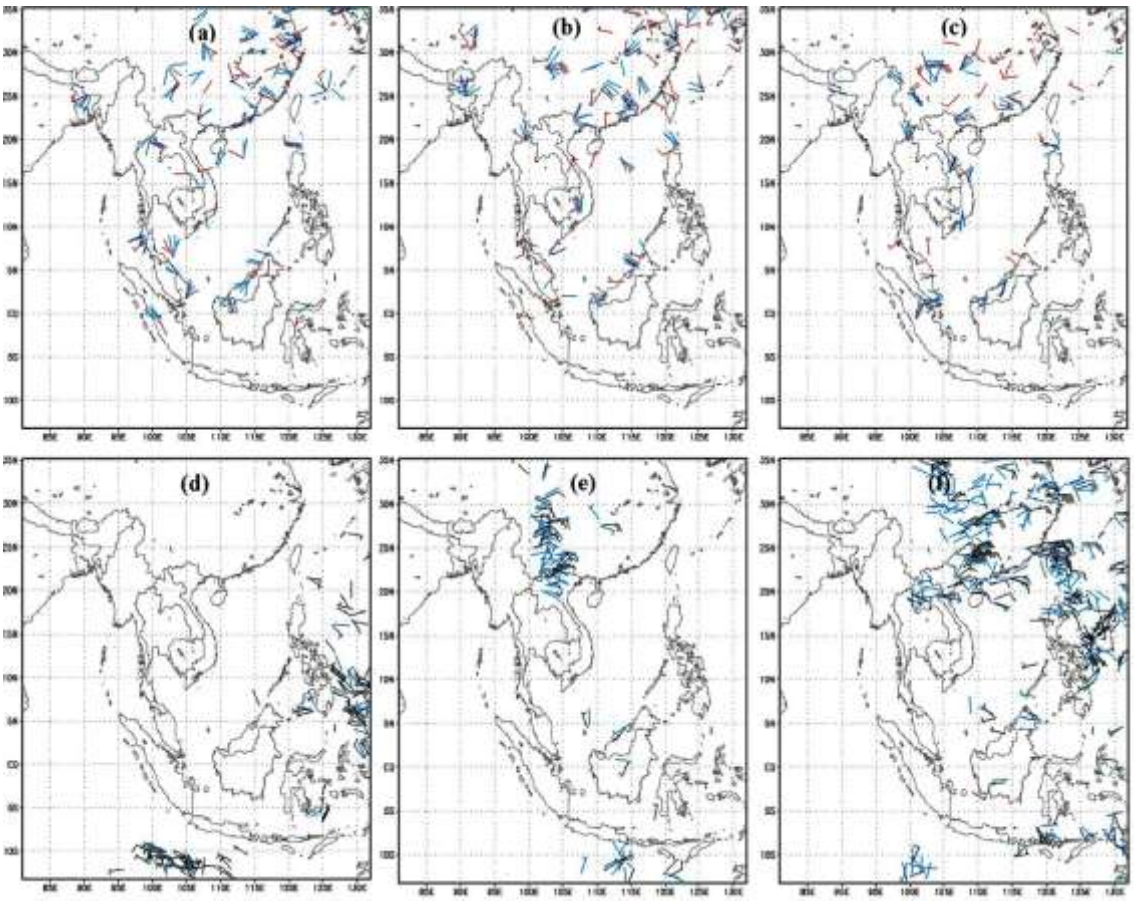

Figure 7. The analytical wind increments (blue barbs) generated by LETKF; the upper-air sounding wind (red barbs); and satellite wind (black barbs) at 750 hPa (a, d), 300 hPa (b, e) and 200 hPa (c, f) levels

Figure 8 shows that corresponding to positions of satellite wind (Figure 8a, b, c) is the location of the analyzed wind vectors (Figure 8d, e, f) on 750 hPa, 300 hPa, and 200 hPa levels at 00 UTC 08 May 2005. Figure 9 is similar to figure 8, which illustrates upper-air sounding wind data is in 850 hPa, 500 hPa, and 300 hPa levels. Figure 9 shows the location where the upper-air sounding wind (Figure 9a, b, c) is the location of the analytical wind vectors (Figure 9d, e, f). This result shows that the Local Ensemble Transform Kalman Filter (LETKF) captures well satellite wind data and upper-air sounding wind data. This result is consistent with previous studies (Kieu et al., 2012, Miyoshi T., and Kunii, M. 2012, Nguyen Duc Phuong, 2013). In addition, because the upper-air sounding wind da-

ta is not much than the satellite wind data (Figure 8 and Figure 9), the analyzed wind vectors in Figure 9 (d, e, f) are less than those in Figure 8.

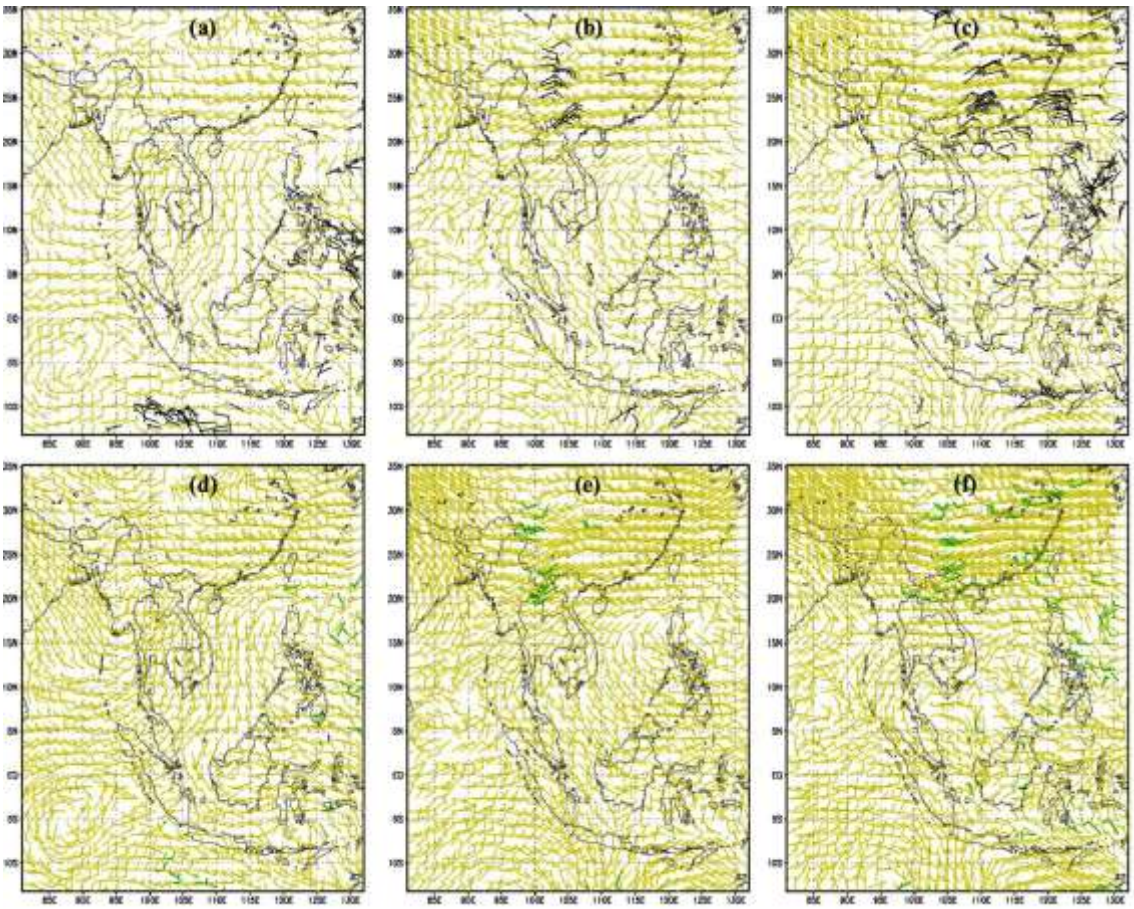

Figure 8. Initial wind field (background field - yellow barbs, b, c, d, e, f); the satellite wind (black barbs - a, b, c) and analysis wind generated by LETKF (d, e and f) (green barbs) at 00 UTC 8 May 2005 at 750 hPa (a, d), 300 hPa (b, e) and 200 hPa (c, f) levels

Figure 10 is a vertical section of the zonal wind component (U) along the longitudinal direction of the Tan Son Hoa meteorological station (106.7°E) in the CTL, CIMSS, RADS, MIX and MPH trials. Figure 10 shows that the west wind zone in the south of 10.8°N (the latitude of the Tan Son Hoa meteorological station) in the data assimilation and multi-physics ensemble experiments are strong and spread to 580 hPa level, whereas CTL test simulates this west wind zone only spread to 600 hPa level (Figure 10a). In the south of 10.8°N, in the layer from 900 hPa level to 60 hPa level, we observe strong west wind zone in the RADS, MPH and CTL trials (6 m/s), while this west wind zone in the CIMSS and MIX trials only reaches 4 m/s. At 200 hPa level, the assimilation data experiments simulate maximum west wind zone which is stronger than that in the CTL and MPH trials, and this west wind zone in the CIMSS test was more than 40 m/s (see Figure 10b). The effect of observation data can be seen more clearly when the wind field is taken in the vertical section of the wind component in the assimilation tests minus the wind field in the vertical section of the wind component in the physical ensemble test (arithmetical difference-AD) (Figure 11). If the AD is posi-

tive (negative) in the western wind area, meaning that the assimilation tests simulate the western wind speed stronger (weaker) than that in the MPH test. Contrarily, if this AD is positive (negative) in the eastern wind area, meaning that the assimilation tests simulate the eastern wind speed weaker (stronger) than that in the MPH test.

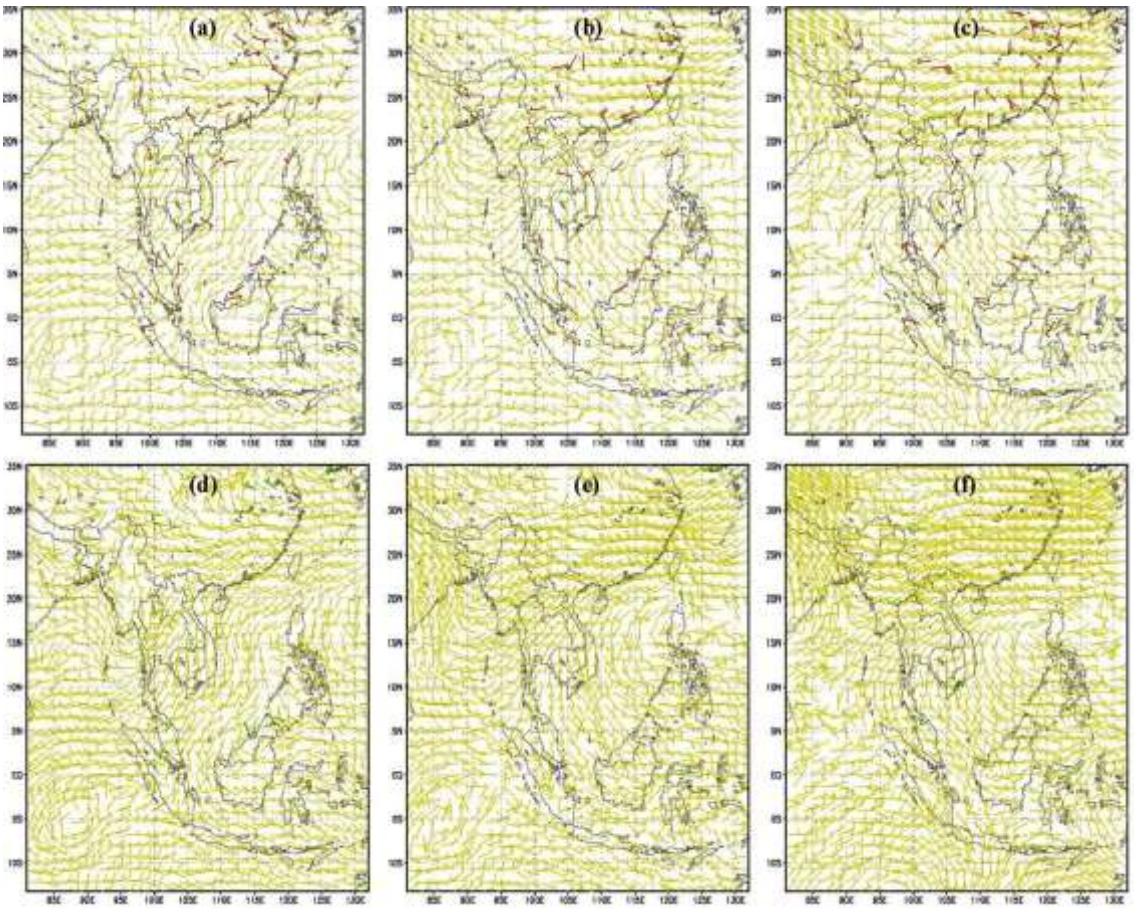

Figure 9. Initial wind field (background field - yellow barbs, b, c, d, e, f); the upper-air sounding wind (red barbs - a, b, c) and analysis wind generated by LETKF (d, e and f) (green barbs) at 00 UTC 8 May 2005 at 850 hPa (a, d), 600 hPa (b, e) and 300 hPa (c, f) levels

In the eastern wind area of the CIMSS test, AD is positive in the layer from 500 hPa to 100 hPa at the southern of the Tan Son Hoa station (Figure 11a). In the western wind area at levels which is under 500 hPa lever, AD is positive at 5°N and from 20°N to 30°N and continued from the surface to 500 hPa level (Figure 11a). The layout of AD in the CIMSS test fits with the distribution of the satellite wind vectors illustrated in Figure 7. In the western wind area of the RADS test at the southern of Tan Son Hoa station, AD is positive and distributed continuously from 900 hPa level to 600 hPa level (Figure 11b). These results indicate that RADS trial simulating the west wind zone in the south of Tan Son Hoa station are stronger than that in the MPH test. The distribution of AD in the RADS test is also consistent with the distribution of upper-air sounding wind data (Figure 8). Meanwhile, AD in the MIX test appears mainly in the 20°N to 30°N area (Figure 11c).

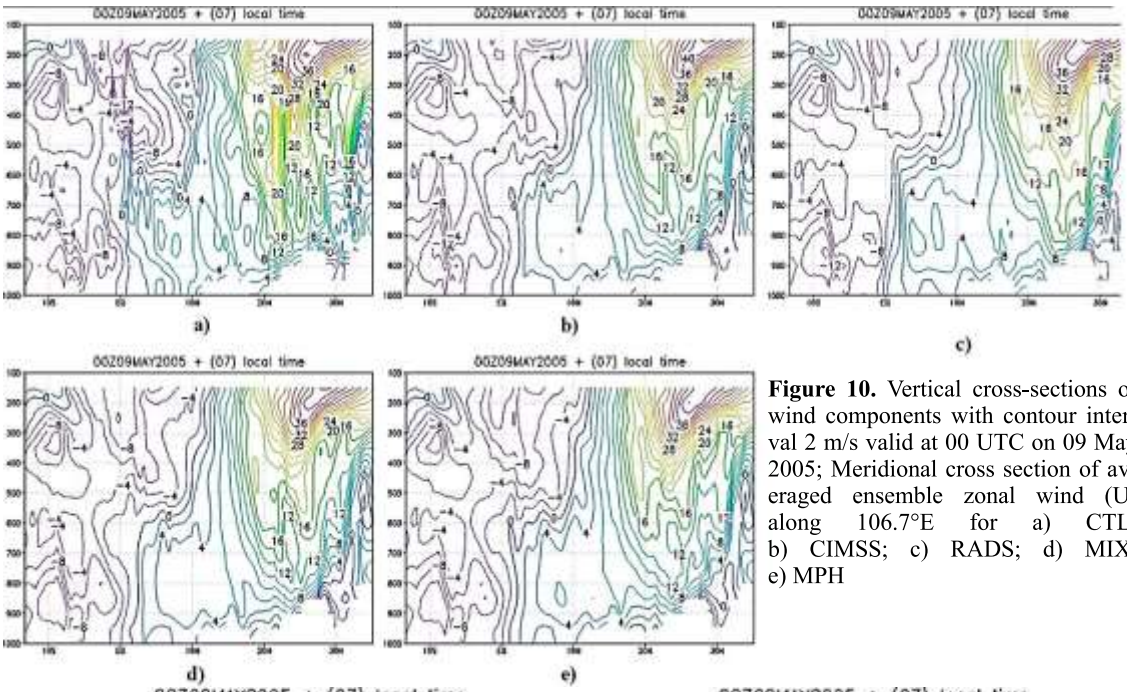

Figure 10. Vertical cross-sections of wind components with contour interval 2 m/s valid at 00 UTC on 09 May 2005; Meridional cross section of averaged ensemble zonal wind (U) along 106.7°E for a) CTL; b) CIMSS; c) RADS; d) MIX; e) MPH

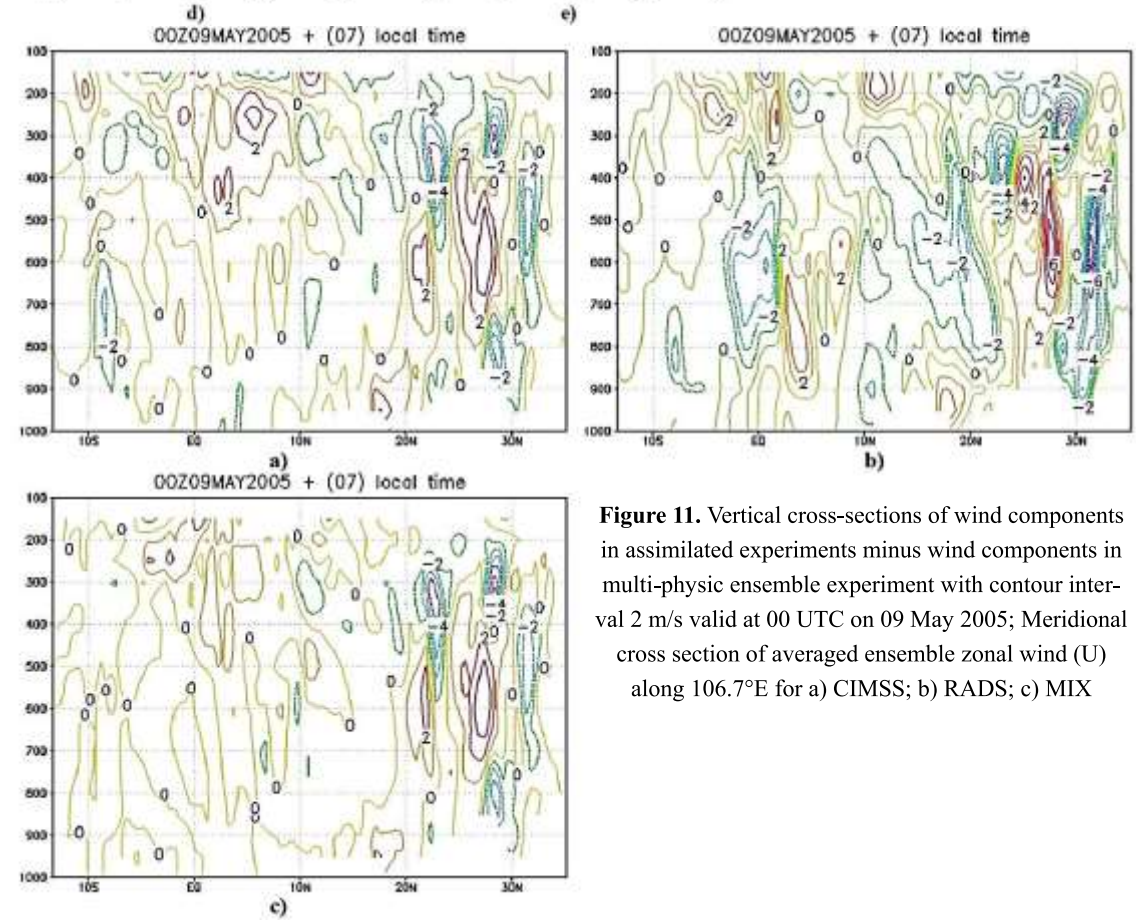

Figure 11. Vertical cross-sections of wind components in assimilated experiments minus wind components in multi-physic ensemble experiment with contour interval 2 m/s valid at 00 UTC on 09 May 2005; Meridional cross section of averaged ensemble zonal wind (U) along 106.7°E for a) CIMSS; b) RADS; c) MIX

For the meridional wind component (Figure 12), the CIMSS, MIX and MPH experiments simulate a strong southern wind zone (8 m/s) and persist from the surface to the 300 hPa level at the west of 106.7°E. The RADS test also simulates this southern wind zone but it is not continuous. The MPH experiment simulates the strongest southern wind (10 m/s) and it prevails continuously from 500 hPa level to 400 hPa level. In the CTL trial, this southern wind zone is weaker than that in other tests. On the 200 hPa level, assimilation and multi-physical ensemble experiments simulate a maximum northern wind zone at the eastern of 120°E (Figures 12b, c, d, e). And the CTL test also simulates a maximum northern wind zone at the western of 120°E. Similar to the zonal wind component, the wind arithmetical difference in the vertical section of the wind component between the assimilation experiments (CIMSS, RADS, and MIX) and the multi-physical ensemble test (MPH) shown in Figure 13. Figure 13 shows the difference in the wind which appears from 1000 hPa level to 100 hPa level in all three CIMSS, RADS and MIX trials.

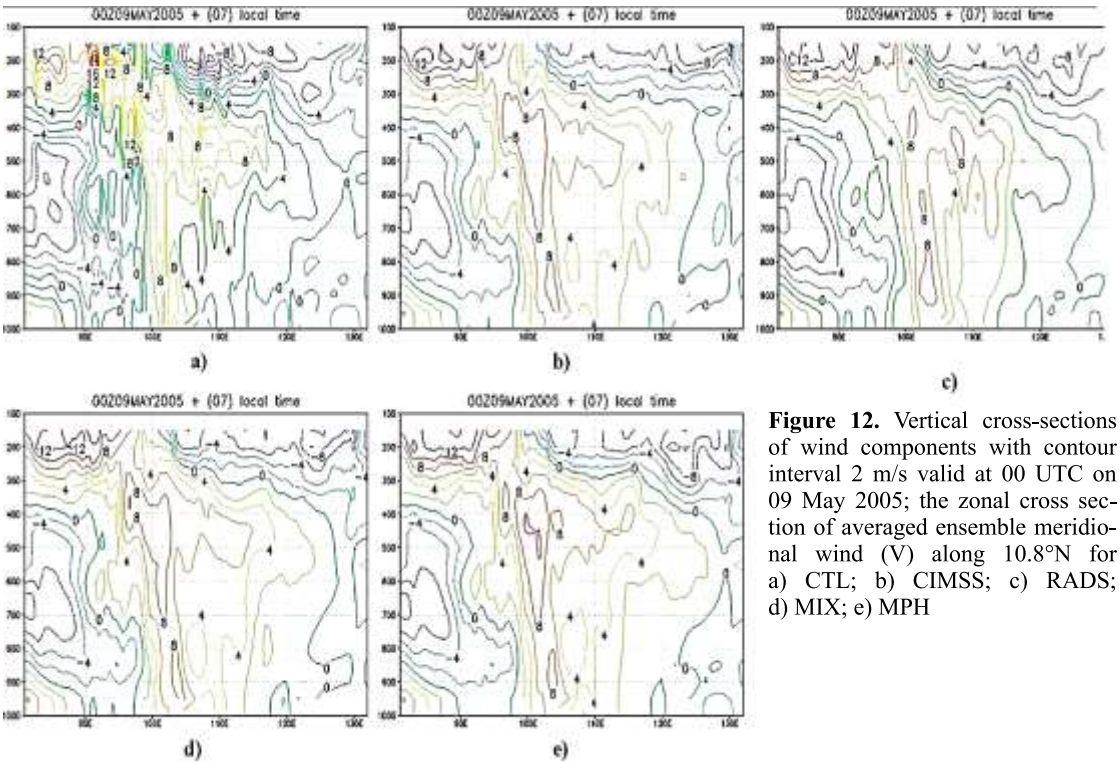

Figure 12. Vertical cross-sections of wind components with contour interval 2 m/s valid at 00 UTC on 09 May 2005; the zonal cross section of averaged ensemble meridional wind (V) along 10.8°N for a) CTL; b) CIMSS; c) RADS; d) MIX; e) MPH

By 10 May 2005, the western wind zone in the south of 10.8°N in the CIMSS, RADS, MIX and MPH trials is narrower than that in the CTL trial (Figure 14). But the maximum western wind center under 600 hPa level in the CIMSS, RADS, MIX and MPH experiments is more extensive than that in the CTL test. The RADS experiment simulates the maximum western wind center (10 m/s) at about 10°N at the 800 hPa level (Figure 14b). The CIMSS, MIX and MPH experiments simulate this maximum western wind zone which translates north and is about 12°N and it has weak intensity (8 m/s). This result can be seen more clearly through arithmetical difference of the wind field between the assimilation tests and the

MPH (AD). Figure 15 shows that AD in the western wind areas in the CIMSS test is positive under the 800 hPa level at the south of 10°N to the equator. These results indicate that the west wind in the CIMSS trial is stronger than the western wind in the MPH trial. The AD in the RADS test is +4 in the layer from 800 hPa level to 600 hPa level (Figure 15b), and the AD in MIX test is much smaller than AD in CIMSS and RADS tests. As such, assimilation tests seem to enhance the west wind zone in the south of 10°N.

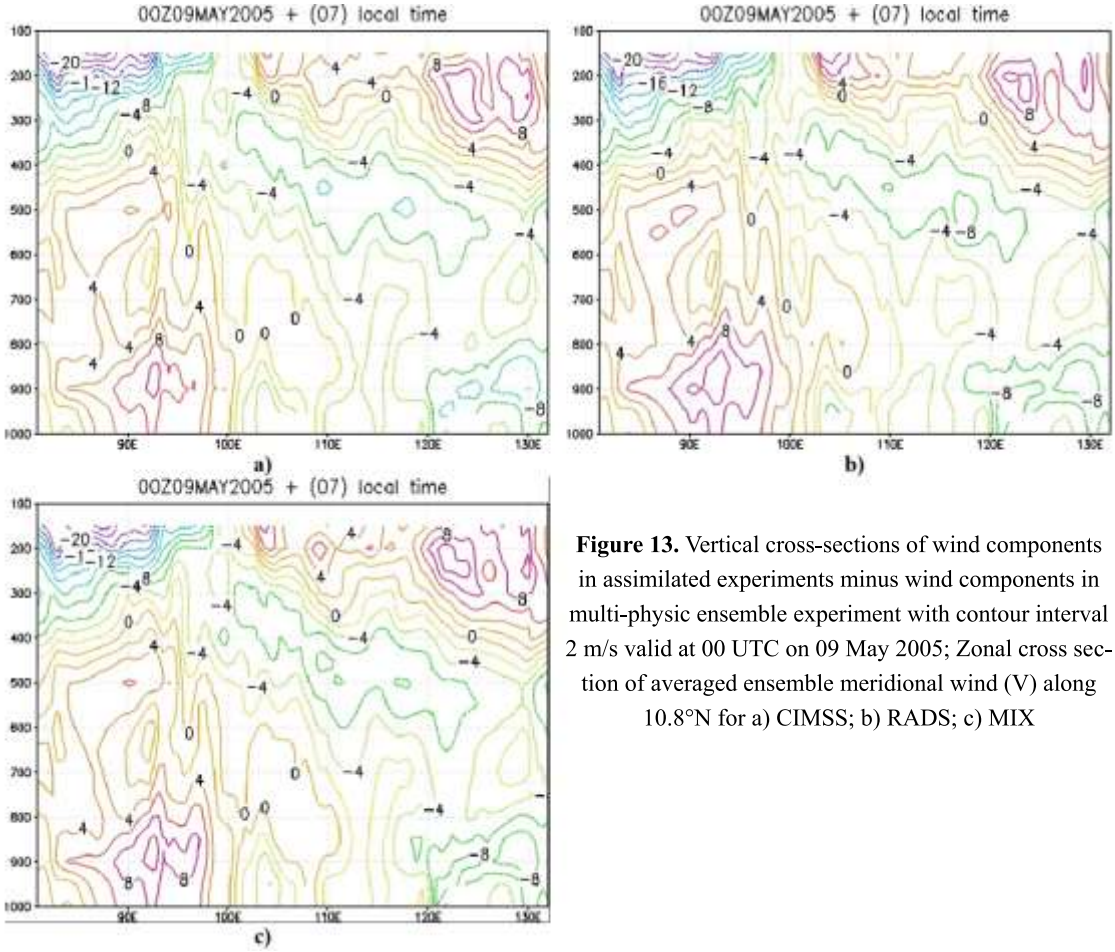

Figure 13. Vertical cross-sections of wind components in assimilated experiments minus wind components in multi-physic ensemble experiment with contour interval 2 m/s valid at 00 UTC on 09 May 2005; Zonal cross section of averaged ensemble meridional wind (V) along 10.8°N for a) CIMSS; b) RADS; c) MIX

For the meridional wind component, the CIMSS, MIX and MPH trials simulate a maximum southern wind zone (10.0 m/s) in the layer from 900 hPa level to 800 hPa level at the west at 106.7°E (Figure 16b, d). Similarly, the RADS test (Figure 16c) simulates the southern wind zone, but its speed is only 8.0 m/s. The CTL test (see Figure 16a) did not simulate this southern wind zone. In addition, when considering arithmetical difference of the meridional wind component between the assimilation experiment and the multi-physical ensemble experiment, will be more clearly show the impact of observed data on the wind field (see Figure 17). From Figure 17, the arithmetical difference in all assimilation experiments occurs in most of the levels. The analysis of the vertical structure of the meridional wind and zonal wind components in the experiments showed that the observed

data changes the vertical structure of the wind components. Particularly 10 May 2005 the observed data seems to strengthen the western wind zone at the south of Tan Son Hoa station. In addition, the results of calculating the root mean square errors (RMSE) of the wind components at Tan Son Hoa station also show the impact of observed data on the wind field (Table 9). In Table 9, the RMSE of the zonal wind component in the CIMSS test is significantly reduced compared to that in the CTL and MPH trials at 24 hours, 24 hours, and 72 hour forecast. For the meridional wind component, the CIMSS trial has RMSE improved at 24 hours and 48 hour forecasts. At the 72 hour forecast, RMSE of the CIMSS trial is larger than that in the CTL and MIX trials, however, the value of RMSE is not significantly different. For RMSE of the meridional wind component in the RADS test is the smallest in compared to that of the remaining trials at the 72 hour forecast.

The impact of observed data on the rain field is examined in the next section.

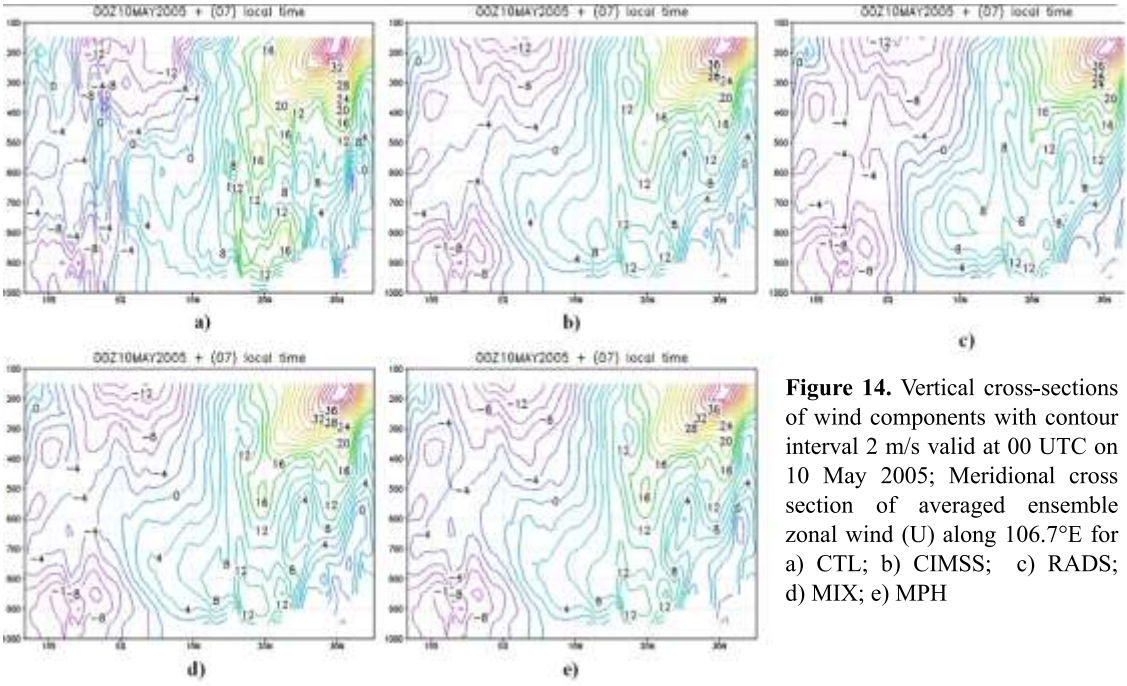

Figure 14. Vertical cross-sections of wind components with contour interval 2 m/s valid at 00 UTC on 10 May 2005; Meridional cross section of averaged ensemble zonal wind (U) along 106.7°E for a) CTL; b) CIMSS; c) RADS; d) MIX; e) MPH

Table 9. Root mean square errors (RMSE) for zonal and meridional wind components at Tan Son Hoa station (10.8°N, 106.7°E) valid at 00 UTC of 08-10 May 2005

Date	RMSE				
(May 2005)	U (m/s)				
	CTL	CIMSS	RADS	MIX	MPH
08	8.19	6.31	7.72	7.51	7.17
09	9.51	8.65	11.63	10.76	10.41
10	9.18	8.71	10.92	10.84	10.61
	V (m/s)				
08	6.28	4.65	5.21	5.04	4.85
09	6.24	6.42	7.17	7.36	7.18
10	6.05	6.26	5.45	6.37	6.15

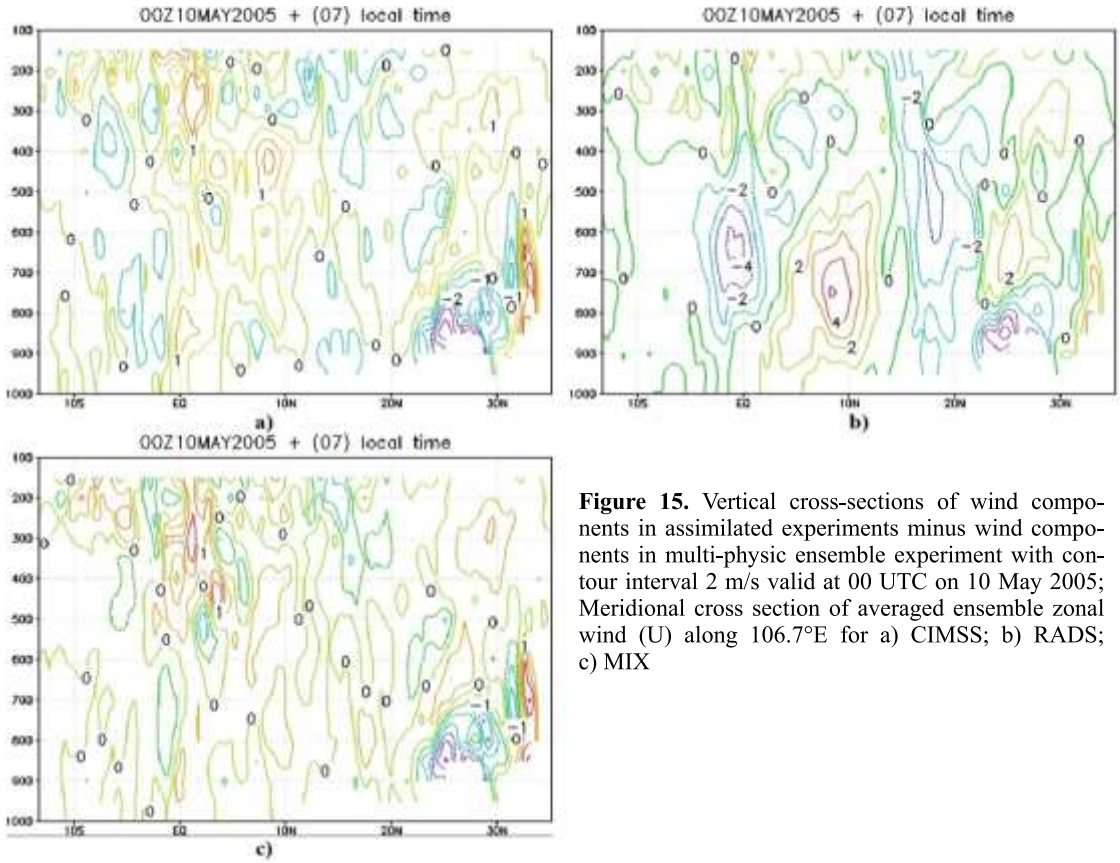

Figure 15. Vertical cross-sections of wind components in assimilated experiments minus wind components in multi-physic ensemble experiment with contour interval 2 m/s valid at 00 UTC on 10 May 2005; Meridional cross section of averaged ensemble zonal wind (U) along 106.7°E for a) CIMSS; b) RADS; c) MIX

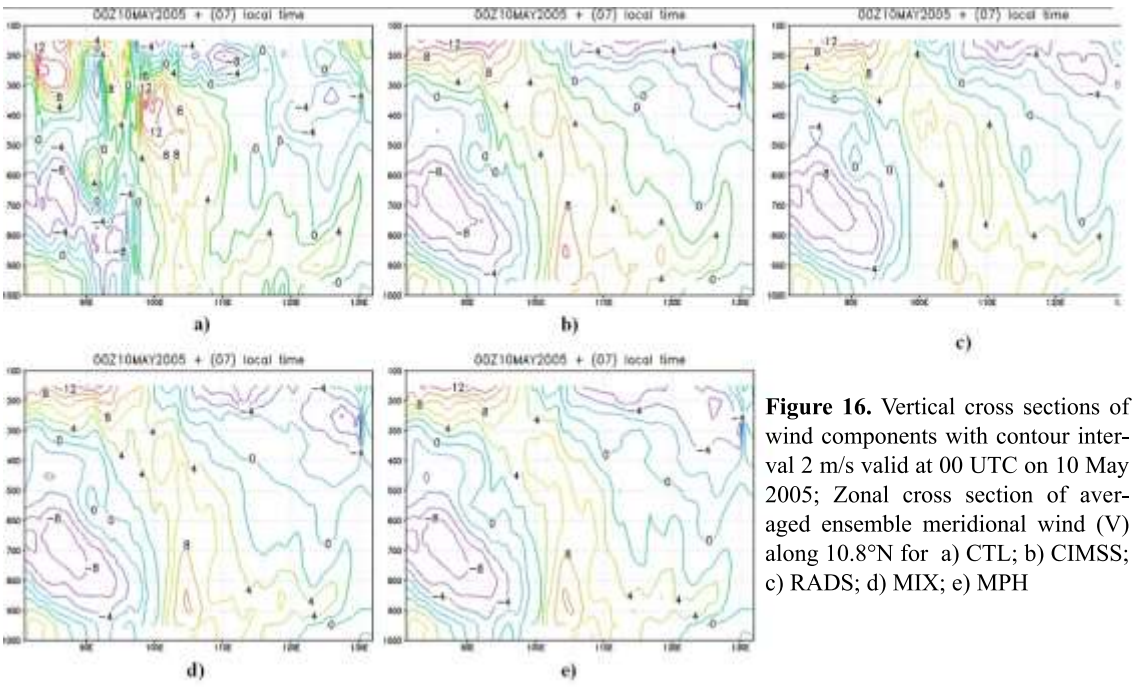

Figure 16. Vertical cross sections of wind components with contour interval 2 m/s valid at 00 UTC on 10 May 2005; Zonal cross section of averaged ensemble meridional wind (V) along 10.8°N for a) CTL; b) CIMSS; c) RADS; d) MIX; e) MPH

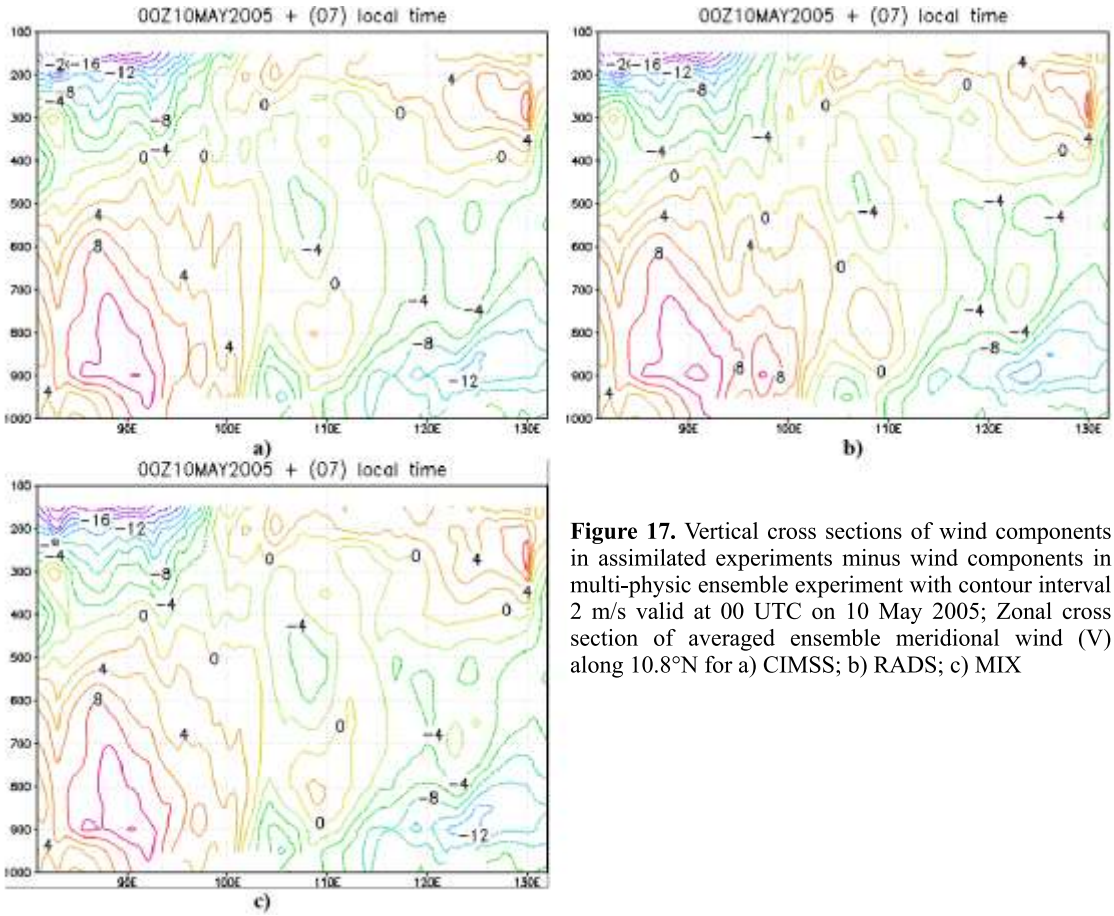

Figure 17. Vertical cross sections of wind components in assimilated experiments minus wind components in multi-physic ensemble experiment with contour interval 2 m/s valid at 00 UTC on 10 May 2005; Zonal cross section of averaged ensemble meridional wind (V) along 10.8°N for a) CIMSS; b) RADS; c) MIX

4.2. Simulation of precipitation measurements

Daily rainfall (unit of mm/day) in assimilation experiments and multi-physical ensemble experiments is the average rainfall of 21 ensemble members and it was extracted to the station. Similarly, precipitation from the CTL test is also extracted from 18 Nam Bo meteorological stations (Table 10). In this section, The effectiveness of assimilation experiments in rainfall simulations is examined from the results of rainfall measurements at 18 stations in Southern Vietnam.

Table 10 shows that the total precipitation in the assimilation data tests and multi-physical ensemble test (CIMSS, RADS, MIX, and MPH) is higher than that observation. Meanwhile, total daily rainfall in the CTL test is lower than the observation on stations where are large daily rainfall. At the 24 hour forecast (Table 10), the CIMSS trial forecasts total daily precipitation at 6 stations closer to observation than that in the CTL test, similarly to the MIX test also predicts total daily precipitation well than that in the CTL test at 6 stations, the RADS test predicts well total rainy days at two stations, and the MPH test predicts well day rainfall at four stations.

At the 48 hour forecast, the RADS test properly predicts rains on six stations (Table 10). The CIMSS and MIX trials forecast total daily precipitation better than that in CTL trial at 5 to 6 stations where have daily rainfall of more than 20mm. The MPH trial predicts better rainfall than CTL test at 3 stations. The re-

sults of this prediction are consistent with the wind analysis results in section 4.1, in which the west wind field at the south of 10.8°N in the RADS test is stronger than the other trials (Figure 10). As a result, the wet is transported from the Bay of Bengal and the Eastern Sea to Southern Vietnam was strengthened. The CIMSS and MIX trials also simulate this west wind field more strongly than in the CTL test, shown at the thickness of the west wind layer which is larger than that in the CTL trial. As a result, the total daily precipitation totals in the assimilation tests are higher than that in the CTL test.

Table 10. Daily rainfall total of observations and simulated models for Day-1, Day-2 and Day-3 valid at 00 UTC of 08-10 May 2005

(mm.day^{-1})

	Day-1						Day-2						Day-3					
Runs / Station	OBS	CTL	CIMSS	RADS	MIX	MPH	OBS	CTL	CIMSS	RADS	MIX	MPH	OBS	CTL	CIMSS	RADS	MIX	MPH
Phuoc Long	0	3	4	12	4	1	0	5	2	4	1	2	32	1	17	27	17	12
Tay Ninh	24	12	18	14	21	23	5	1	7	15	7	7	15	6	26	24	27	24
Tan Son Hoa	0	5	6	8	5	2	16	1	1	3	2	1	28	1	6	17	8	4
Ba Tri	7	2	8	10	6	4	0	2	4	5	3	3	0	0	12	12	13	9
Vung Tau	0	4	15	21	15	14	0	6	10	22	9	9	36	2	25	34	31	21
Cang Long	0	4	18	15	16	14	11	2	1	8	1	1	27	3	16	23	20	14
Moc Hoa	34	9	16	12	17	18	0	2	2	7	2	1	9	6	19	23	18	17
Rach Gia	27	2	2	3	2	3	14	1	1	8	1	2	5	4	13	18	12	10
Cao Lanh	0	6	11	8	10	20	21	2	5	11	3	4	5	9	16	25	15	17
Chau Doc	0	4	5	7	4	6	17	3	2	4	1	1	9	11	30	18	18	16
Can Tho	2	7	19	14	22	23	1	5	7	11	9	10	0	7	19	22	20	19
My Tho	0	7	15	16	17	17	0	4	8	13	7	10	29	5	26	30	27	23
Soc Trang	3	6	24	21	21	19	13	6	8	14	8	11	0	7	23	21	23	19
Ca Mau	9	3	4	5	3	5	47	1	5	10	6	8	15	4	10	12	11	9
Bac Lieu	0	5	14	12	13	14	5	3	2	9	2	3	0	3	10	19	12	7
Phu Quoc	7	2	16	16	15	14	22	4	35	56	35	31	20	22	46	44	82	38
Tho Chu	0	2	7	16	7	6	3	11	20	28	24	27	5	13	30	23	31	23
Con Dao	19	5	16	18	14	14	9	1	6	9	5	5	22	1	10	15	11	7

At the 72 hour forecast (Table 10), the RADS test properly predicts total rainfall over 8 stations where have total rainfall greater than 20mm. The CIMSS trial forecasts rainfall near observation at 3 stations. The MIX test forecasts positively rainfall on 5 stations. While the MPH test forecasts effectively rainfall on 2 stations. However, the RADS, MIX and MPH tests predicted total rainfall which is a large difference than observation in the other stations.

For the rainy event during the monsoon onset period of 2008 and 2009 years, the daily rainfall on the stations is calculated similarly to that rainy event of the 2005 year. Table 11 is the total daily rainfall in 2008 (28/4 to 30/4) which is forecasted at 00 UTC 28 April 2008. The CTL, MPH tests and assimilation tests forecast rainfall which appears over Southern Vietnam. However, the CIMSS trial forecasts rainfall near observation at some stations (Ca Mau Station (28/4), Cang Long, Rach Gia, Can Tho, Phu Quoc and Tho Chu stations (29/4)). By 30 April, Cang Long Station has total rainfall of up to 100mm/day, but most of the experiments do not predict this total rainfall. The MIX trial also forecasts rainfall near observation in some stations such as Rach Gia, Moc Hoa and Phu Quoc (29/4); The RADS test showed some positive rainfall results in Phu Quoc station (29/4), Con Dao station (30/4); Meanwhile, the MPH test forecasts well rainfall in Can Tho, Phu Quoc, Con Dao stations on April 29 and April 30 (Table 11).

Table 11. Daily rainfall total of observations and simulated models for Day-1, Day-2 and Day-3 valid at 00 UTC of 28-30 April 2008

(mm.day^{-1})

Runs / Stations	Day-1						Day-2						Day-3					
	OBS	CTL	CIMSS	RADS	MIX	MPH	OBS	CTL	CIMSS	RADS	MIX	MPH	OBS	CTL	CIMSS	RADS	MIX	MPH
Phuoc Long	0	8	31	15	21	14	50	13	18	9	9	12	0	5	29	8	18	14
Tay Ninh	0	8	17	8	13	13	1	16	9	4	12	8	0	71	28	17	25	19
Tan Son Hoa	2	8	8	6	7	12	42	9	13	4	20	8	0	1	17	4	22	25
Ba Tri	0	20	20	21	16	19	1	35	26	27	22	20	54	8	10	22	18	15
Vung Tau	0	11	8	2	4	10	0	1	11	8	36	11	20	6	8	6	15	12
Cang Long	0	11	10	11	6	10	27	28	25	16	13	22	100	12	14	28	20	16
Moc Hoa	0	9	9	4	10	10	16	12	12	3	14	8	6	51	18	17	18	20
Rach Gia	0	29	26	25	25	28	45	90	32	24	41	21	2	31	30	49	36	23
Cao Lanh	0	15	16	8	13	12	25	13	13	13	16	13	9	24	20	29	20	17
Chau Doc	0	15	38	14	23	25	5	14	27	10	29	32	41	16	17	42	18	11
Can Tho	0	13	15	8	11	9	31	21	19	12	14	23	4	10	22	25	34	31
My Tho	0	9	11	6	12	8	1	3	15	8	23	19	42	1	20	10	23	12
Soc Trang	0	28	15	17	11	19	2	17	32	18	16	15	50	31	27	29	22	22
Ca Mau	36	22	39	23	33	23	10	35	29	43	22	22	5	115	26	65	54	44
Bac Lieu	1	21	13	8	10	14	2	33	25	32	20	17	11	105	35	54	28	28
Phu Quoc	0	30	37	38	42	30	56	20	35	46	35	30	0	30	20	33	19	19
Tho Chu	0	34	22	45	26	31	21	14	30	81	56	38	0	16	13	33	21	18
Con Dao	0	8	19	27	17	5	2	16	29	27	26	25	41	35	36	48	30	36

For 2009, rainfall event is forecasted from April 29 to May 1 at 00 UTC 29 April 2009. During this rainy event, the RADS and the MPH tests forecast positively rainfall on a number of stations belong Southern Vietnam area at 72 hour forecast (Table 12). The CIMSS and MIX trials improved the rainfall forecasting quality at 48 hour forecast (Table 12). The CTL test gives positive predictive results for all forecasting periods. This result may be explained by the fact that the observed rainfall on April 29, April 30 and May 1 at stations is very small. On the other hand, the boundary conditions and initial conditions of the CTL test is GFS data. While GFS data is often lower than that observed. Therefore the results of the rain forecast in the CTL is more favorable than those of the assimilation and multi-physical ensemble tests.

Table 12. Daily rainfall total of observation and simulated model for Day-1, Day-2 and Day-3 valid at 00 UTC of 29-30 April and 01 May 2009

mm.day^{-1})

Runs / Stations	Day-1						Day-2						Day-3					
	OBS	CTL	CIMSS	RADS	MIX	MPH	OBS	CTL	CIMSS	RADS	MIX	MPH	OBS	CTL	CIMSS	RADS	MIX	MPH
Phuoc Long	31	16	4	12	4	1	15	11	19	23	18	16	13	17	28	28	36	34
Tay Ninh	34	8	18	14	21	23	10	7	9	37	16	12	52	17	23	43	15	17
Tan Son Hoa	7	7	6	8	5	2	8	4	9	12	11	7	0	23	20	24	22	24
Ba Tri	6	0	8	10	6	4	0	0	6	21	4	3	57	6	10	14	9	13
Vung Tau	11	16	15	21	15	14	1	8	6	14	24	18	26	9	20	25	18	25
Cang Long	0	9	18	15	16	14	0	5	3	18	13	9	25	5	5	17	3	7
Moc Hoa	2	9	16	12	17	18	0	5	4	23	14	5	12	12	19	17	10	14
Rach Gia	4	4	2	3	2	3	0	0	3	7	4	3	4	4	9	9	4	5
Cao Lanh	22	4	11	8	10	20	1	5	3	16	10	5	2	10	8	11	6	10
Chau Doc	7	9	5	7	4	6	1	4	2	14	7	5	2	3	9	9	4	6
Can Tho	0	8	19	14	22	23	0	5	4	22	8	3	9	5	7	12	5	9
My Tho	0	17	15	16	17	17	0	4	11	23	21	12	29	14	24	19	16	23
Soc Trang	8	11	24	21	21	19	2	4	2	18	11	8	12	10	1	13	2	4
Ca Mau	0	3	4	5	3	5	0	0	2	25	4	2	3	0	1	6	1	1
Bac Lieu	8	5	14	12	13	14	0	6	1	14	8	6	0	1	1	5	1	1
Phu Quoc	0	5	16	16	15	14	2	0	3	14	5	4	0	2	9	14	7	9
Tho Chu	0	1	7	16	7	6	0	0	1	11	2	2	3	0	4	12	3	3
Con Dao	0	1	16	18	14	14	0	1	4	23	10	2	12	5	2	14	3	3

The next section we use the statistical index and phase prediction index is calculated based on the average of 21 ensemble members of the assimilation and multi-physical ensemble tests in 18 trials and observation data at 18 Nam Bo meteorological stations. The total of statistical samples is 36.

4.3. Evaluating the prediction ability of the model

4.3.1. Statistical indeces

Table 13 shows the statistical indices for the trials. At the 24 hour forecast, the two RADS and MIX trials have ME negative, meaning that the rainfall forecasts is lower than that observation, while the three CIMSS, MPH, and CTL trials have positive ME. However, the ME only reflects the average deviation of the predicted value compared to the observation, so only based on ME cannot evaluate the predictability of the tests. The MAE indicates error average value. The MAE in the RADS and MIX trials are smaller than that in the remaining trials (Table 13). This result shows that the assimilation of the upper-air sounding data and mixed data in the input field of the model have a positive effect on the predicted rainfall at 24 hours period. The RMSE index is similar to the MAE. But RMSE can amplify the error when the predicted value deviates much than the observation. And, therefore the evaluating results is not objective. If RMSE is much larger than MAE, there is a large difference between predicted and observed values. Table 13 shows the RMSE values ranged from 14 mm to 17 mm. The arithmetical difference of the RMSE and MAE values ranged from 2 mm to 7 mm. Two RADS and MIX tests had a large arithmetical difference of the RMSE and MAE values (4 mm to 7 mm). This result indicates a greater disparity between predicted precipitation and observed rainfall in these tests compared to the remaining trials (CTL, CIMSS, and MPH).

Table 13. Statistical evaluation indexs at the station

Index	Day-1					Day-2					Day-3				
	CTL	CIMSS	RADS	MIX	MPH	CTL	CIMSS	RADS	MIX	MPH	CTL	CIMSS	RADS	MIX	MPH
ME	1.48	4.16	-8.60	-8.66	3.42	1.63	3.00	7.71	4.73	3.75	-0.59	3.77	7.07	4.51	0.55
MAE	10.96	12.86	9.89	8.66	12.00	13.11	12.48	15.30	14.00	13.02	16.65	15.61	16.68	16.72	15.66
RMSE	14.73	16.69	16.76	15.55	15.85	19.65	16.99	19.06	18.40	17.40	24.75	21.27	21.14	22.77	21.34
RMSE-MAE	2.66	2.24	6.87	6.07	3.85	5.74	3.87	3.75	4.07	4.38	2.64	2.02	4.46	4.83	5.67

At the 48 hour forecast, most of the experiments predicted higher rainfall than observed, as shown in positive ME index (Table 13). In addition, the MAE and RMSE of the CIMSS test are the smallest, meaning that the CIMSS test's capable of predicting rainfall is better than the remaining trials. Thus, the Local Ensemble Transform Kalman Filter assimilates the satellite wind data to improve the rainfall forecasting quality at the 48 hour range.

At 72 hour forecast, statistical indices are also calculated similarly to the 24 hour and 48 hour forecasts (Table 13). The results show that the CIMSS trial had significantly improved in MAE compared to MAE in CTL and MPH tests. While RADS and MIX experiments do not significantly improve in MAE compared to MAE of CTL and MPH tests. The arithmetical difference between the RMSE and MAE values in the CIMSS trial is smaller than others at 24 hours and 72 hour forecasts.

As a result, the analysis of statistical indicators for 2005, 2008 and 2009 trials showed that satellite wind data assimilation test predicted good rainfall at 48 hours and 72 hours ranges. And upper-air sounding data and mix data assimilation tests have a positive impact on rainfall forecasts at 24 hours. In addition, multi-physical ensemble prediction (MPH test) has partly contributed to the improvement of rainfall forecasts at 24 and 48 hours. The results of the MPH trial are consistent with the findings of Cong Thanh et

al. (2015).

4.3.2. Phase prediction index

In this section, we assess the model's predictability through frequency indices between predicted "yes" and observed "yes" (BIAS); the correct predicting percentage of the actual (POD); The predicted false is the percent of total occurrences (FAR) and the indicator of fit between the forecasting area and the observed area (ETS). Table 14 compares Phase prediction indexes of CTL test, data assimilation tests (CIMSS, RADS, and MIX) and multi-physical ensemble test (MPH).

Table 14. The phase indexs in the tests applied to the 6mm/24 hour rainfall threshold

Range	Day-1				Day-2				Day-3			
Index	BIAS	POD	FAR	ETS	BIAS	POD	FAR	ETS	BIAS	POD	FAR	ETS
CTL	1.48	0.46	0.68	0.23	1.53	0.45	0.70	0.21	1.23	0.37	0.69	0.20
MPH	1.72	0.47	0.72	0.20	1.86	0.64	0.65	0.29	1.79	0.62	0.65	0.29
CIMSS	2.72	0.51	0.75	0.20	1.93	0.63	0.67	0.27	1.98	0.67	0.66	0.29
RADS	2.17	0.37	0.82	0.13	2.10	0.61	0.70	0.24	2.01	0.69	0.66	0.29
MIX	1.90	0.50	0.74	0.20	2.04	0.67	0.67	0.28	2.07	0.71	0.65	0.30

At the 24 hour forecast, the predicted trials had BIAS index higher than 1, meaning that the model predicts higher precipitation than the observation. This result is consistent with the analysis of ME in section 4.3.1. In terms of value, the CTL test which has BIAS value close to 1 compared to other tests. Considering the POD and ETS indicators: These two indicators assess the predictive power of the model, while the POD evaluates the probability of occurrence events, the ETS index measures the probability of the events does not happen. The 24 hour forecast shows a significant improvement in the CIMSS trial (POD = 0.51), while the RADS test had the smallest ETS, meaning that the ability to assess the non-occurrence of the event in upper-air sounding data assimilation test is low. The FAR of the RADS test is higher than other tests, which means that the misses forecast rate in the RADS test is high (82%).

Within the 48 hour forecast, the trials has BIAS higher 1, meaning that the model predict rain in the non-rain cases. With the POD index, the MIX test had the largest POD (67%), while CTL test had the lowest POD (45%). Similarly, the MIX, CIMSS and MPH tests have good predictive skills, whereas CTL and RADS tests have bad predictive skills. The value of the FAR index indicates that the RADS and CTL tests had percentage of false predictions which is higher than the remain tests (MPH, CIMSS and MIX).

At the 72 hour forecast, the BIAS is similar to this index at the 24 hour and 24 hour. Specifically, the CTL test has the smallest BIAS index and greater than 1, while the assimilation experiments have the BIAS index larger than 2. For this prediction range, the POD index in the assimilation tests (CIMSS, RADS, and MIX) is comparable and higher than that in the two control trials (CTL and MPH). Whereas the FAR index in assimilation and multi-physical ensemble tests are similar and lower than that in the CTL test. Similar to the ETS index in CIMSS, RADS, MIX and MPH trials is higher than that in the CTL trial.

In addition, we base on the average value and volatility of the index to give the overall ranking of the forecast level of the tests in order from best (1) to lowest quality (5) for each indicator (Table 15). At the same time, we combined with the evaluation from the statistical indicators (section 4.3.1) to make the following remarks: MIX and CIMSS trials are likely to improve the rainfall forecast quality in the Southern monsoon onset period at the 48 hour and 72 hour ranges, for the RADS test improved the 72 hour rainfall forecast. In ad-

dition, the multi-physics ensemble test predicts well precipitation at 24 hours and 48 hours ranges. This result is consistent with the previous study (Cong Thanh et al., 2015).

Table 15. Overall rating of the predictive power of experiments

Range	Day-1				Day-2				Day-3			
Index	BIAS	POD	FAR	ETS	BIAS	POD	FAR	ETS	BIAS	POD	FAR	ETS
CTL	1	4	1	1	1	5	3	5	1	5	3	3
MPH	2	3	2	2	2	2	1	1	2	4	1	2
CIMSS	5	1	4	2	3	3	2	3	3	3	2	2
RADS	4	5	5	3	5	4	3	4	4	2	2	2
MIX	3	2	3	2	4	1	2	2	5	1	1	1

5. Conclusions

The study tested 18 cases of precipitation forecast during the Vietnam Southern monsoon outbreaks in 2005, 2008 and 2009 years. Each of the cases performs 85 predictions, including the data assimilation tests (3 trials and 63 predictions), multi-physical ensemble test (21 predictions) and single predictive test. The results of the data assimilation and multi-physical ensemble tests were calculated to average on an ensemble of 21 members and were compared with the observation data at 18 Vietnam Southern meteorological stations. From the above analysis results can be drawn as follows:

From the results of the 2005 rainfall forecast (May 8th to May 10th) show that the Local Ensemble Transform Kalman Filter (LETKF) has capable of capturing satellite wind and upper-air sounding data. This result is consistent with previous studies on LETKF (Kieu et al., 2012; Miyoshi T. and Kunii M., 2012; Phuong Nguyen Duc 2013), so that the initial analysis field with higher resolution is better than analytical data of the GFS global model with low resolution.

The assimilation of satellite wind data, upper-air sounding data, and mixed data into the input field of the WRF model by LETKF has partly improved the input field of the model. This result is shown by the change of the wind field on levels (section 4.1). Thus, the assimilation of data into the input field of the model restores the observation data that has been lost in the interpolation of GFS global modeling data in the high resolution of the regional model. As a result, data assimilation experiments successfully simulate the wind field and positively influenced the rainfall forecasting skills of the model.

The rain forecast results of the experiments show that rain forecasts in the satellite data assimilation and mixed data test are significantly improved at 48 hours and 72 hour forecast, upper-air sounding data assimilation test produces satisfactory rainfall forecasts at 72 hour forecast, and multi-physics ensemble trials predicted good rainfall at 24 and 48 hours. Therefore, it is possible to confirm that satellite wind data and upper-air sounding data are reliable sources of data. Therefore, we can assimilate satellite wind data and upper-air sounding data into the input field of the model in meteorological forecasting problems.

References

Bui Minh Tuan, Nguyen Minh Truong, 2013. Determining the onset indexes for the summer monsoon over southern Vietnam using numerical model with reanalysis data. VNU Journal of Science, 29(1S), 187-195.

Charney J.G., 1955. The use of the primitive equations of motion in numerical prediction, Tellus, 7, 22.

Cong Thanh, Tran Tan Tien, Nguyen Tien Toan, 2015. Assessing prediction of rainfall over Quang Ngai area of Vietnam from 1 to 2 day terms. VNU Journal of Science, 31(3S), 231-237.

Courtier P., Talagrand O., 1987. Variational assimilation of meteorological observations with the adjoint vor-

ticity equations, Part II, Numerical results. Quart. J. Roy. Meteor. Soc., 113, 1329.

Daley R., 1991. Atmospheric data analysis. Cambridge University Press, Cambridge.

Elementi M., Marsigli C., Paccagnella T., 2005. High resolution forecast of heavy precipitation with Lokal Modell: analysis of two case studies in the Alpine area. Natural Hazards and Earth System Sciences, 5, 593-602.

Fasullo J. and Webster P.J., 2003. A hydrological definition of India monsoon onset and withdrawal. J. Climate, 16, 3200-3211.

Haltiner G.J., Williams R.T., 1982. Numerical prediction and dynamic meteorology, John Wiley and Sons, New York.

Hamill T.M., Whitaker J.S., Snyder C., 2001. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. Mon. Wea. Rev., 129, 2776.

He J., Yu J., Shen X., and Gao H., 2004. Research on mechanism and variability of East Asia monsoon. J. Trop. Meteo, 20(5), 449-459.

Hoang Duc Cuong, 2008. Experimental study on heavy rain forecast in Vietnam using MM5 model. A report on the Ministerial-level research projects on science and technology, 105p.

Houtekamer P.L., Mitchell H.L., Pellerin G., Buehner M., Charron M., Spacek L., Hansen B., 2005. Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. Mon. Wea. Rev., 133, 604.

Houtekamer P.L., Mitchell H.L., 2005. Ensemble Kalman filtering, Quart. J. Roy. Meteor. Soc., 131C, 3269-3289.

Hunt B.R., Kostelich E., Szunyogh I., 2007. Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter. Physica D., 230, 112-126.

Kalnay E., 2003. Atmospheric modeling, data assimilation and predictability. Cambridge University Press, 181.

Kalnay et al., 2008. A local ensemble transform Kalman filter data assimilation system for the NCEP global model. Tellus A, 60(1), 113-130.

Kato T., Aranami K., 2009. Formation Factors of 2004 Niigata-Fukushima and Fukui Heavy Rainfalls and Problems in the Predictions using a Cloud-Resolving Model. SOLA. 10, doi:10.2151/sola.

Kieu C.Q., 2010. Estimation of Model Error in the Kalman Filter by Perturbed Forcing. VNU Journal of Science, Natural Sciences and Technology, 26(3S), 310-316.

Kieu C.Q., 2011. Overview of the Ensemble Kalman Filter and Its Application to the Weather Research and Forecasting (WRF) model. VNU Journal of Science, Natural Sciences and Technology, 27(1S), 17-28.

Kieu C.Q., Truong N.M., Mai H.T., and Ngo Duc T., 2012. Sensitivity of the Track and Intensity Forecasts of Typhoon Megi (2010) to Satellite-Derived Atmosphere Motion Vectors with the Ensenble Kalman filter. J. Atmos. Oceanic Technol., 29, 1794-1810.

Kieu Thi Xin, 2005. Study on large-scale rainfall forecast by modern technology for flood prevention in Vietnam. State-level independent scientific and technological briefing report, 121-151.

Kieu Thi Xin, Vu Thanh Hang, Le Duc, Nguyen Manh Linh, 2013. Climate simulation in Vietnam using regional climate nonhydrostatic NHRCM and hydrostatic RegCM models. Vietnam National University, Hanoi. Journal of Natural sciences and technology, 29(2S), 243-25.

Krishnamurti T.N., Bounoa L., 1996. An introduction to numerical weather prediction techniques. CRC Press, Boca Raton, FA.

Lau K.M., Yang S., 1997. Climatology and interannual variability of the Southeast Asian summer monsoon. Adv. Atmos. Sci., 14,141-162.

Li C., Qu X., 1999. Characteristics of Atmospheric Circulation Associated with Summer monsoon onset in the South China Sea. Onset and Evolution of the South China Sea Monsoon and Its Interaction with the Ocean. Ding Yihui, and Li Chongyin, Eds, Chinese Meteorological Press, Beijing, 200-209.

Lin N., Smith J.A., Villarini G., Marchok T.P., Baeck M.L., 2010. Modeling Extreme Rainfall, Winds, and Surge from Hurricane Isabel, 25. Doi: 10.1175/2010WAF2222349.

Lu J., Zhang Q., Tao S., and Ju J., 2006. The onset and advance of the Asian summer monsoon. Chinese Science Bulletin, 51(1), 80-88.

Matsumoto J., 1997. Seasonal transition of summer rainy season over Indochina and adjacent monsoon region. Adv. Atmos. Sci., 14, 231-245.

Miyoshi T., and Kunii M., 2012. The Local Ensenble Transform Kalman Filter with the Weather Rearch and Forecasting Model: Experiments with Real Observation. Pure Appl. Geophysic, 169(3), 321-333.

Miyoshi T., Yamane S., 2007. Local ensemble transform Kalman filtering with an AGCM at a T159/L48 resolution. Mon. Wea. Rev., 135, 3841-3861.

Nguyen Khanh Van, Tong Phuc Tuan, Vuong Van Vu, Nguyen Manh Ha, 2013. The heavy rain differences based on topo-geographical analyse at Coastal Central Region, from Thanh Hoa to Khanh Hoa. Vietnam Journal of Earth Sciences, 35, 301-309.

Nguyen Minh Truong, Bui Minh Tuan, 2013. A case study on summer monsoon onset prediction for southern Vietnam in 2012 using the RAMS model. VNU Journal of Science, 29(1S), 179-186.

Phillips N.A., 1960b. Numerical weather prediction. Adv. Computers, 1, 43-91, Kalnay 2004.

Phillips N., 1960a. On the problem of the initial data for the primitive equations, Tellus, 12, 121126.

Phuong Nguyen Duc, 2013. Experiment on combinatorial Kalman filtering method for WRF model to forecast heavy rain in central region in Vietnam. The Third International MAHASRI/HyARC Workshop on Asian Monsoon and Water Cycle, 28-30 August 2013, Da Nang, Viet Nam, 217-224.

Richardson L.F., 1922. Weather prediction by numerical process. Cambridge University Press, Cambridge. Reprinted by Dover (1965, New York).

Routray, Mohanty U.C., Niyogi D., Rizvi S.R., Osuri K.K., 2008. First application of 3DVAR-WRF data assimilation for mesoscale simulation of heavy rainfall events over Indian Monsoon region. Journal of the Royal Meteorological Society, 1555.

Schumacher, R. S., C. A. Davis, 2010. Ensemble-based Forecast Uncertainty Analysis of Diverse Heavy Rainfall Events, 25. Doi: 10.1175/2010WAF2222378.

Snyder C., Zhang F., 2003. Assimilation of simulated Doppler radar observations with an Ensemble Kalman filter. Mon. Wea. Rev., 131, 1663.

Szunyogh I., Kostelich E.J., Gyarmati G., Kalnay E., Hunt B.R., Ott E., Satterfield E., Yorke J.A., 2008. A local ensemble transform Kalman filter data assimilation system for the NCEP global model. Tellus A., 60, 113-130.

Tanaka M., 1992. Intraseasonal oscillation and the onset and retreat dates of the summer monsoon east, southeast Asia and the western Pacific region using GMS high cloud amount data. J. Meteorol. Soc. Japan, 70, 613-628.

Tan Tien Tran, Nguyen Thi Thanh, 2011. The MODIS satellite data assimilation in the WRF model to forecast rainfall in the central region. VNU Journal of Science, Natural Sciences and Technology, 27(3S), 90-95.

Tao S., Chen L., 1987. A review of recent research on East summer monsoon in China, Monsoon Meteorology. C. P. Changand T. N. Krishramurti, Eds, Oxford University Press, Oxford, 60-92.

Tippett M.K., Anderson J.L., Bishop C.H., Hamill T.M., Whitaker J.S., 2003. Ensemble square root filters. Mon. Wea. Rev., 131, 1485.

Thuy Kieu Thi, Giam Nguyen Minh, Dung Dang Van, 2013. Using WRF model to forecast heavy rainfall events on September 2012 in Dong Nai River Basin. The Third International MAHASRI/HyARC Workshop on Asian Monsoon and Water Cycle, 28-30 August 2013, Da Nang, Viet Nam, 185-200.

Xavier, Chandrasekar, Singh R. and Simon B., 2006. The impact of assimilation of MODIS data for the prediction of a tropical low-pressure system over India using a mesoscale model. International Journal of Remote Sensing 27(20), 4655-4676. https://doi.org/10.1080/01431160500207302.

Wang B., 2003. Atmosphere-warm ocean interaction and its impacts on Asian-Australian monsoon variation. J. Climate, 16(8), 1195-1211.

Wang B. and Wu R., 1997. Peculiar temporal structure of the South China Sea summer monsoon. J. Climate., 15, 386-396.

Wang L., He J., and Guan Z., 2004. Characteristic of convective activities over Asian Australian "land-bridge" areas and it spossible factors. Act a Meteorologic a Sinica, 18, 441-454.

Wang, B., and Z. Fan, 1999. Choice of South Asian Summer Monsoon Indices. Bull. Amer. Meteor. Sci., 80, 629-638.

Webster P.J., Magana V.O., Palmer T.N., Shukla J., Tomas R.A., Yanai M., Yasunari T., 1998. Monsoons: Processes, predictability, and teprospects for prediction, J. Geophys. Res., 103, 14451-14510.

Wilks Daniel S., 1997. Statistical Methods in the Atmospheric Scienes. Ithaca New York., 59, 255.

Whitaker J.S., Hamill T.M., 2002. Ensemble data assimilation without perturbed observations. Mon. Wea. Rev., 130, 1913.

Wu G., Zhang Y., 1998. Tibetan plateau forcing and the timing of the monsoon onset over South Asia and the South China Sea. Mon.Wea.Rev., 126, 913-927.

Zhang Z., Chan J.C.L., and Ding Y., 2004. Characteristics, evolution and mechanisms of the summer monsoon onset over Southeast Asia. J.Climatology, 24, 1461-1482.

http://weather.uwyo.edu/upperair/sounding.html and http://tropic.ssec.wisc.edu/archive/