Impact of surface data assimilation using an ensemble Kalman filter technique on the improvement of weather prediction

N J Trilaksono1, M Taqyya2,3, N Dewani4,5, I D G A Junnaedhi1, E Riawan1, and R Fitriani4

1 Atmospheric Science Research Group, Faculty of Earth Sciences and Technology, Institut Teknologi Bandung, Jalan Ganesha 10, Bandung, ID
2 Meteorology Program, Institut Teknologi Bandung, Jalan Ganesha 10, Bandung, ID
3 International Graduate Program for Advanced Science, Tohoku University, 6-3 Aramaki Aza Aoba Sendai, JP
4 Earth Sciences Program, Institut Teknologi Bandung, Jalan Ganesha 10, Bandung, ID
5 UPC Renewables Indonesia, Sunset Road No. 29 C, Kuta Bali, ID
E-mail: jpatiani@meteo.itb.ac.id

Abstract. Initial condition data is known to be an important factor, which contributes to the accuracy of a weather prediction. One of an attempt to improve the initial condition is by applying data assimilation method. Ensemble Kalman Filter (EnKF) is a sophisticated data assimilation method that has shown great promise for atmospheric data assimilation in a way that the method uses flow-dependent error statistics estimated from relatively short period of prediction. Data assimilation of surface observations using an EnKF technique is evaluated for the potential improvement of a regional nonhydrostatic model. We use Weather Research and Forecasting - Advanced Research WRF (WRF-ARW) model to implement the EnKF data assimilation system.

The EnKF system combined with the WRF-ARW model on a 1-km horizontal grid spacing over Karimunjawa Island is cycled over a field course period from 14–18 August 2015. The analysis ensemble is generated every 6 h by assimilating surface observations from three different locations of automatic weather stations (AWSs). By assimilating the surface observation data, the models produce colder temperature over the sea near Karimunjawa Island compare to those without assimilation process. Thus, the assimilation data give a stronger and more defined sea breeze circulation because the circulation generated by temperature difference between the sea and land.

1. Introduction

Sea breeze is a coastal local wind that blows from sea to land, caused by the temperature difference between the sea and the adjacent land. The sea breeze has been recognized for centuries throughout the world and extensively studied using observational data and model simulation (e.g., [1][2]). In the tropics, which is characterized by a prominent feature of diurnal cycle of precipitation, the sea breeze is associated with the maximum peak of precipitation over land in the afternoon [3][4]. Although the important role of the sea breeze has been relatively well understood, there is still difficulty in forecasting the strength and correct phase of the
phenomenon. Many of numerical studies of the diurnal cycle related to the sea breeze have reported shortcomings in reproducing the phase and amplitude of the phenomenon [5][6].

On the other hands, initial condition data for model simulation is one of the important aspect, which contribute to the accuracy of the model output. However, it also contributes to the source of error in the model simulation, whereas this problem has already recognized very long time ago by Bjerknes [7]. To provide better initial condition, data assimilation method can be employed to obtain the best estimate of state of the atmosphere [8]. Meanwhile, ensemble-based data assimilation system, such as the ensemble Kalman filter (EnKF) has shown to give reliable results for atmospheric data assimilation because the system use flow-dependent error statistics estimated from short-term ensemble forecasts [9].

This study describes the performance of an EnKF system over the sea breeze circulation over a small island. The primary goal of this study is to evaluate the use of automatic weather station data in improving daily weather prediction. This paper proceeds as follows. Section 2 describes the modeling and data assimilation system, while section 3 provides the results of the experiments. A summary and discussion are provided in section 4.

2. Experimental Design
In this study, ensemble analyses are generated every 6 h by cycling an EnKF from 06 UTC 15 to 00 UTC 16 August 2015. Figure 1 shows the computational domain, which consists of three nested domains. The outer domain has $192 \times 130$ grids with a grid spacing of 20 km, centered at $6.38^\circ$S, $110.15^\circ$E on a Mercator projection. While the two inner domains are nested with the grid spacing ratio of 3, thus they have 3 and 1 km grid spacings, respectively. All ensemble members are advanced in time using the Advanced Research Weather Research and Forecasting model (ARW-WRF, hereafter WRF; [10]) version 3.5.1 with 27 vertical levels. The WRF subgrid-scale parameterizations and their parameters follow the work by [11]. The National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS) with the horizontal resolution of $0.25^\circ \times 0.25^\circ$ and the time interval of 3 h are used for both model initial and boundary data.

Ensemble initial and lateral boundary conditions are generated by drawing random perturbations from the NCEP error covariances contained in the WRF VAR system. Observation data from three Automatic Weather Stations (AWSs) sited in the Karimunjawa Island, — site 1: (5.885764°S, 110.4457°E), site 2: (5.838771°S, 110.421465°E), and site 3: (5.772914°S, 110.484846°E)—are assimilated using the Data Assimilation Research Testbed (DART; [12]), which is an implementation of the ensemble adjustment Kalman filter. Only three variables from each AWS are included in the assimilation process, namely u-wind and v-wind component and temperature.

![Figure 1. Computational domain. Symbols d01, d02 and d03 denote the domain number.](image-url)
3. Results

3.1. Spatial comparison of cloud top temperature between satellite data and model output

A qualitative analysis is done by comparing the forecast without assimilation against the Himawari 8 satellite data to verify the numerical result. A comparison between the Himawari 8 satellite IR1 channel data and a derived variable, which mimic Multi-functional Transport Satellite 2 (MTSAT2) brightness temperature, from the model output shows a comparable scale of cloud feature. Generally, the model is able to reproduce the small scale feature of cloud particularly over the land area, regardless the periode of study that tend to be clear sky over southern part of Indonesia. However, we notice artificial features near the edge of computational domain, which associated with boundary effect problem.

As mention in the previous section, the initial and boundary condition are from GFS $0.25^\circ \times 0.25^\circ$ and the outer domain has 9 km grid spacing. However, the model output shows reproducibility of small scale features similar to the satellite data, which has about 5 km of horizontal resolution. This result implies that the model is useful to represent the atmospheric condition to some extent.

3.2. Statistical evaluation of assimilation experiment

Statistical analyses are performed to evaluate the impact of data assimilation to the improvement of prediction, particularly the wind component. Figure 2 shows comparison of RMSE of u-wind and v-wind component for with and without data assimilation. Note that site 4 is another station sited at Jepara, which located in the mainland of Java Island. There is a significant improvement of RMSE for both u-wind and v-wind component compare to those without assimilation. The smaller RMSE for the assimilation result denote that the assimilation process reduce the error of the model compare to the observation. However, a higher RMSE for data assimilation experiment exists at site 2 for v-wind component.

Correlation analysis between wind from the model output and the observation also shows an increase of correlation coefficient, which implies an improvement of model output. In particular, the significant increase of the correlation coefficient are for v-wind component. Although, the correlation coefficient for u-wind component also increase, the correlation is in reverse relation at site 1–3.

![Figure 2](image.png)

**Figure 2.** Comparison of RMSE between control run - observation and assimilation - observation for a) u-wind component and b) v-wind component. Blue color denotes control run - observation and red color denotes assimilation - observation.
3.3. Spatiotemporal distribution of surface temperature

The sea breeze circulation strongly relates with the temperature difference between the sea and adjacent land as many studies point out (see [1][2]). Therefore, it is important to examine the surface temperature parameter from the model output to determine the sea breeze existence. In Figure 3, we present comparison of spatial distribution of the surface temperature from the model output with and without assimilation. For the EnKF system, we only run five ensemble members due to limited of computational resources.

A relatively large difference of surface temperature appears near Karimunjawa Island between with and without assimilation. All of the ensemble member of EnKF system show a colder temperature in the sea surrounding the Karimunjawa Island compare to the result without assimilation. Although careful investigation are required, we still can expect that there will be stronger sea breeze circulation in the ensemble member of EnKF than in the control run (without assimilation) due to stronger temperature gradient between the sea and the land in Karimunjawa Island.

![Figure 3.](image)

Figure 3. Comparison of surface temperature (shaded; in unit K) and horizontal wind (vector; in unit m s$^{-1}$) from model output a) without and b–f) with assimilation (member 1–5) for a particular time (15 August 2015 15:00 local time).

4. Summary and Conclusions

This study presents the impact of AWS data assimilation to the daily prediction that include local circulation over a small island. A comparison between the Himawari 8 satellite IR1 channel data and brightness temperature from the model output shows a good agreement of small scale features, particularly over the Java Island. Statistical analyses show that the data assimilation can improve the prediction as represented by decrease of RMSE values. EnKF data assimilation system produce colder surface temperature over the sea surrounding Karimunjawa Island, which can be expected produce stronger sea breeze circulation over the island.
Acknowledgments

We thank Glen Romine (NCAR) for his guidance to run WRF/DART system. This work is supported by funds provided by Institut Teknologi Bandung Research and Innovation Program 2016 under Grants FITB.PN-6-10-2016.

References

[1] Miller S T K, Keim B D, Talbot R W and Mao H 2003 Sea breeze: structure, forecasting, and impacts Rev. Geophys. 41 1011
[2] Hadi T W, Horinouchi T, Tsuda T, Hashiguchi H and Fukao S 2002 Mon. Wea. Rev 130 2153–66
[3] Sato T, Miura H, Satoh M, Takayabu Y N and Wang Y 2009 J. Climate 22 4809–26
[4] Zhou L and Wang Y 2006 J. Geoph. Res. 111 D17104
[5] Teo C-K, Koh T-Y, Lo J C-F and Bhatt B C 2011 J. Climate 24 4662–75
[6] Koo M-S and Hong S-Y 2010 J. Geoph. Res. 115 D05105
[7] Bjerknes V 1911 Dynamic Meteorology and Hydrography. Part II. Kinematics (New York: Carnegie Institute, Gibson Bros.)
[8] Kalnay E 2003 Atmospheric Modeling, Data Assimilation and Predictability (Cambridge: Cambridge Univ. Press) p 341
[9] Torn R D 2010 Mon. Wea. Rev. 138 4375–92
[10] Skamarock W C, Klemp J B, Dudhia J, Gill D O, Barker D M, Duda M G, Huang X-Y, Wang W and Powers J G 2008 A Description of the Advanced Research WRF Version 3 (Boulder, Colorado: National Center for Atmospheric Research) p 125
[11] Gustari I, Hadi T W, Hadi S and Renggono F 2012 Jurnal Meteorologi dan Geofisika 13 119–30
[12] Anderson J L, Hoar T, Raeder K, Liu H, Collins N, Torn R and Arellano A 2009 Bull. Amer. Meteor. Soc. 90 1283–96