MINISTRY OF RESOURCES AND ENVIRONMENT
VIET NAM INSTITUTE
OF METEOROLOGY, HYDROLOGY AND CLIMATE CHANGE

Nguyen Hai Dong

ASSIMILATION RESEARCH OF SATELLITE DATA FOR COMMUNITY MULTISCALE AIR QUALITY MODEL (CMAQ) IN HANOI AREA

Major: Resource and environment management
Code: 9850101

SUMMARY OF DOCTORAL DISSERTATION ON RESOURCE AND ENVIRONMENT MANAGEMENT

Hanoi - 2021

The project was completed at:

Viet Nam Institute of Meteorology, Hydrology and Climate Change

Science instructor:

1. Assoc.Prof.Dr. Doan Ha Phong
2. Dr. Le Ngoc Cau

PhD dissertation reviewer 1:

PhD dissertation reviewer 2:

The thesis will be defended in front of the Institute's Thesis Judging Committee meeting at: Institute of Meteorology, Hydrology and Climate Change

at date month year 2021

The dissertation can be found at:

- National Library of Vietnam;
- Library of Institute of Meteorology, Hydrology and Climate Change.

INTRODUCTION

1. Regarding Rationale of the Research

Air is the amount of gas that always surrounds us, air is colorless, odorless, tasteless and it is a determining factor for human life as well as all living things on earth.

Air pollution is when there is a major change in the composition of the air or when there is an appearance of strange gases that make the air no longer clean, have odors, reduce visibility, and cause climate change as well as disease in humans and animals. Air pollution happens when the air contains harmful components such as gases, suspended dust, smoke, odors exceeding a certain threshold [3].

There are many standards for assessing air quality, however, the concentration of suspended particles in the air, especially $PM_{2.5}$ and PM10, has been widely accepted to assess the quality of air. Therefore, the term air quality will refer to ambient $PM_{2.5}$ concentrations in the remainder of this study.

To assess air quality, many methods as well as technical solutions have been developed such as measurement method using equipment at ground-based monitoring stations, remote sensing method (through sensors installed on satellites) and modeling methods (using mathematical models).

For the measurement method using equipment at ground-based monitoring stations, air pollutants provide a qualitative and quantitative way of concentration and deposition. However, they can only describe air quality at specific locations and times without giving direction on determining the cause of air pollution. Remote sensing method using sensors placed on satellites is used to assess air quality on a large scale at the same time. Despite that, this method does not currently meet the requirements of accuracy, frequency of information provision, and predictability.

Community Multiscale Air Quality (CMAQ) is a modeling system capable of simulating complex atmospheric processes affecting variability, propagation, and deposition.

However, for the modeling method, the input factor of the model plays a very important role. In fact, currently, there is no method capable of providing complete input data of the CMAQ model, which requires integration from many different data sources. These are also difficult issues and a key point to be solved for the problem of air quality modeling.

On that basis, the research topic "***Assimilation Research of Satellite Data for Community Multiscale Air Quality Model (CMAQ) in Hanoi Area***" was selected.

In order to carry out this study, several questions were proposed as follows:

- In which module of the model data assimilation for CMAQ model is performed?
- What kind of satellite data meets the requirements for assimilation purposes and how is the assimilation of satellite data for the CMAQ model carried out?
- How to evaluate the results of the model after the satellite data has been assimilated? Effectiveness of satellite data assimilation in air pollution monitoring?

2. Objectives of the Research

- Research and apply The Local Ensemble Transform Kalman Filter (LETKF) to assimilate AOD data from MODIS satellite, improve the accuracy of estimating $PM_{2.5}$ concentration in the air for Hanoi area.

- Research and propose the process of assimilating AOD satellite data using the WRFDA module to serve the assessment of air quality in accordance with the conditions of Vietnam.

3. Research content

- Overview, collection, analysis and evaluation of documents: topography, meteorology, remote sensing data, mathematical model, WRF-CMAQ model system, mathematical basis and data assimilation algorithm, research documents related to the content of the research;
- Analyzing and processing remote sensing images to calculate the AOD value as input for the WRF-CMAQ model system;
- Structure, modules of the WRF-CMAQ model system, settings, input data sources, implementation steps and simulation results;
- Simulation on the basis of before and after assimilating AOD satellite data, calculate $PM_{2.5}$ concentrations on WRF-CMAQ model system;
- Display, analyze and evaluate the results of $PM_{2.5}$ concentrations on simulation results, compare and correlate with monitoring data at monitoring stations, causes and movement of $PM_{2.5}$ polluting atmosphere in the research area.

4. Object and scope of the research

- ***Scope of space***: Hanoi and surrounding areas.
- ***Time range:*** The years of 2015, 2017 and 2019.
- ***Scope of research content***: Assimilation of AOD satellite data for WRF-CMAQ model system.
- ***Research object:*** Fine dust ($PM_{2.5}$) concentration; WRF-CMAQ model system; Kalman filter; MODIS remote sensing data; Aerosol Optical Depth (AOD).

5. Research Methods

The research methodology used in this study include:

- Consultants method: Consult with experts in the field of air pollution measurement and monitoring;

- Methods of processing remote sensing images: Perform processing, directly calculate AOD data, dust concentration on MODIS remote sensing images;

- Modeling method: Satellite data assimilation technique for CMAQ model;

- Statistical method, multi-time analysis: Analyze the evolution of $PM_{2.5}$ concentration from the results of the model over simulation time.

6. Scientific and Practical Significance

- ***Scientific significance:*** The results of the thesis provide a scientific basis, a new method using satellite data to build a set of input data, a process of assimilation of satellite data for the WRF-CMAQ model system in order to create a tool for assessing air quality in Hanoi in particular and Vietnam in general.

- Practical significance: The results of the dissertation can be applied to the assessment of air quality, assisting managers in the field of monitoring and identifying sources of polluting emissions, thereby considering, deciding on activities to reduce air pollution and develop socio-economic.

7. Defensive Arguments

- Argument 1: The AOD digitalization protection process uses the WRFDA module to serve the forecasting and assessment of air quality in accordance with the conditions of Vietnam.

- ***Argument 2:*** Assimilation of AOD data from MODIS satellite improves estimation accuracy, air quality forecast for Hanoi area is appropriate.

8. New contributions of The Research

Establishing a new method in air quality assessment through the comprehensive CMAQ model for air pollution monitoring in Hanoi and expanding to all provinces and cities nationwide and is especially useful for areas where there is no ground monitoring station.

9. Dissertation Layout

In addition to the introduction; references; appendix, the dissertation structure insisted of 03 chapters:

Chapter 1. Overview of Assimilation Methods.

Chapter 2. Mathematical Basis and Research Methods.

Chapter 3. Results of Assimilated Satellite Data for the WRF-CMAQ Model System.

Conclusions and Recommendations

CHAPTER 1. OVERVIEW OF ASSIMILATION METHODS

1.1. Overview of environmental monitoring methods

Currently, there are many monitoring methods for air pollution monitoring that are being applied and deployed in practice.

a. Method of determination by monitoring equipment

The method of determining basic parameters in the air by monitoring equipment at monitoring stations to measure the concentration of substances in the air is being widely used today, especially in urban areas. towns and industrial zones.

Studying air pollution by remote sensing image data is based on the scattering characteristics of electromagnetic waves on polluted dust particles in the atmosphere.

Scientists have proposed many different methods to determine air pollution from satellite image data such as: Carnahan et al., 1984; Tanre et al., 1988; Sifakis and Deschamps, 1992; Retalis et al., 1998; Nuno Grosso et al., 2007; Retalis and Sifakis, 2009; Chu et al., 2003.

b. Modeling method

The modeling method uses mathematical models to describe the process of pollutant diffusion as well as to calculate with the help of computers to calculate the concentrations of material and chemical components. in the air.

1.2. Concept, overview of assimilation methods

In 2003, Eugenia Kalnay introduced the concept "Data assimilation is the best combination of different sources of information, data, observation data, background data, a priori information or statistics, to estimate, calculate the state of a system, a model equation".

Data assimilation in the earth sciences was early applied to meteorology as Lorenc, 1986; Daley, 1991; Kalnay, 2003; Evensen, 2009;. Lahoz et al, 2010; Fedorov, 1989; Daley, 1991.

Optimal statistical interpolation approach was applied by Fedorov, 1989; Daley, 1991; Ghil and Malanotte-Rizzoli, 1991; Penenko and Obraztsov, 1976; Le Dimet and Talagrand, 1986; Talagrand, 1987; Rabier, 2000; Lorenc, 2003; Xavier, 2006; Routeray, 2008; Kalnay et al., 2008.

1.3. Overview of assimilation in the field of meteorological data and atmospheric chemistry

The increasing number of satellite data through the retrieval of atmospheric air composition signatures makes assimilation of satellite data in air quality simulation an increasingly possible method in order to obtain more accurate analysis results and initial conditions for air quality forecasting. Significant examples are Kursinski et al., 1997; Rocken et al., 1997; Xavier, 2006; Routeray, 2008; Rakesh, 2009.

The satellite data has great potential for use in chemical assimilation, i.e. data assimilation for CTM chemical models such as Carmichael et al., 2008; Zhao and Wang, 2009; Lamsal et al., 2008; Ionov et al., 2006; Boersma et al., 2008; Huijnen et al., 2010; Boersma et al., 2008.

In Vietnam, Kieu Thi Xin and Le Duc, 2003 applied the 3D-VAR assimilation model to the HRM model. Kieu Thi Xin, 2005 also used a two-dimensional variational method to analyze soil moisture from 2 meter monitoring temperature for the HRM model. Another authors, Tran Tan Tien and Nguyen Thi Thanh, 2011 have studied the assimilation of MODIS satellite data in the WRF model to forecast heavy rain in the Central region. Prof. Phan Van Tan and Nguyen Le Dung, 2009 have tested and applied the WRF-VAR system combined with the vortex initialization scheme to forecast the storm trajectory in the East Sea.

Author Kieu Quoc Chanh, 2011 has reviewed the combined Kalman filter assimilation system and its application to the WRF weather forecasting model.

Recently, assimilation of data by the combined method is being developed and applied by major meteorological centers around the world, especially the combined Kalman filter method.

By assimilation of data, forecasting errors due to initial conditions can be mitigated leading to better forecasting results. The more accurate the initial condition estimation, the better the forecast quality will be. Data assimilation has become an important method in the forecasting industry.

Sub-conclusion of chapter 1

There is no doubt that combining multiple data sources for the purpose of generating the best initial estimate from a given predictive state (or background forecast) of the atmosphere as input to the WRF-CMAQ simulation model, improving the quality of air quality forecasting results is an inevitable trend of forecasting.

Data from earth observation satellites are increasingly high accuracy and high frequency. The advantage of remote sensing methods is the ability to provide data on a large scale, from local to national, even global scale; especially the price is quite low, even free of charge. The results of previous studies have verified that aerosol optical depth data from satellites has an impact on improving the accuracy of estimates of airborne concentrations. Determining the method of processing AOD data plays an important role in assimilation, creating an input data source for the WRF-CMAQ model system.

Assimilation of meteorological and atmospheric chemistry data has also been shown to significantly improve the results of forecasts. The mathematical-based assimilation technique of the combined Kalman filter has been applied to the WRF model system to assimilate satellite data to predict the atmospheric state as well as the air quality have been done and proven effective by many scientists. However, previous research results have not mentioned the combination of WRF model system and CMAQ model system to evaluate air quality based on Kalman filter for AOD satellite data assimilation.

Therefore, the process of assimilating AOD satellite data for the WRF-CMAQ model system was developed to establish a new method for comprehensive air quality assessment on a large scale for monitoring air pollution in the Hanoi area, extending over the territory of Vietnam, is an issue that needs to be studied.

CHAPTER 2. MATHEMATICS BASIS AND RESEARCH METHODS

2.1. Data assimilation technique

According to Talagrand and Kalnay, the best estimate of the state of the atmosphere obtained from the statistical combination of the first guess about the atmosphere and the observations is the state of the analysis. In order to obtain an optimal estimate, statistical information about the errors in the observations is required.

Another method to obtain an optimal estimate is Kalman filtering. Kalman filtering is an algorithm that provides estimates of some unknown variable based on observed measurements over time. Kalman filter has been proving its usefulness in various applications.

2.1.1. Algorithm of the extended Kalman filter

The extended Kalman filter (EKF) is used for nonlinear applications. The algorithm of the extended Kalman filter consists of two steps: the "*prediction step*" of the forecast state and its covariance matrix, and the "*analysis step*" that updates the analysis state and covariance respectively, which are summarized as follows:

1. Input

Forecast state $x^a(t_0) = x_0$

and background error covariance matrix $\mathbf{P}^a(t_0) = \mathbf{P}_0$

2. Loop i = 1, 2, ...

• Prediction step:

$$x^f(t_i) = M_{i-1}\left[x^a(t_{i-1})\right]$$

$$\mathbf{P}^f(t_i) = \mathbf{L}_{i-1}\mathbf{P}^a(t_{i-1})\mathbf{L}_{i-1}^T + \mathbf{Q}_{i-1}$$

• Analysis step:

$$\mathbf{K}_i = \mathbf{P}^f(t_i)\mathbf{H}_i^T\left[\mathbf{R}_i + \mathbf{H}_i\mathbf{P}^f(t_i)\mathbf{H}^T\right]^{-1}$$

$$x^a(t_i) = x^f(t_i) + \mathbf{K}_i\left[y_i^0 - H\left[x^f(t_i)\right]\right]$$

$$\mathbf{P}^a(t_i) = \left[\mathbf{I} - \mathbf{K}_i\mathbf{H}_i\right]\mathbf{P}^f(t_i)$$

An excellent characteristic of the extended Kalman filter is that even if a system starts with a poor initial guess about the state of the atmosphere, the EKF will provide the best linear estimate of the state and its covariance. However, there is a limitation in the application that the error propagation is approximated by the tangent linear model between the two analysis steps.

2.1.2. Algorithm of the combinatorial Kalman filter

Since the possibility of developing a tangent model and integrating the covariance error matrix over time with the tangent model is impractical in predictive models, the Kalman filter has been improved to be applicable to professional problems. A variant based on the stochastic Monte-Carlo integral, whereby a set of inputs based on the probability distribution as well as the error value of the analytic field at each point in time is generated around a given integral field value. The given analysis has been developed and is called The Local Ensemble Transform Kalman Filter (LETKF). This is also the method that will be selected and included in the WRFDA assimilation module later in this study.

Basically, LETKF is a method that, at each mesh node, selects a model neighborhood of a given size. Then use the background combinatorial noise matrix to transform from local space to combinatorial space. This greatly reduces the matrix computation because the combinatorial space is often much smaller than the local space. As a result, matrix operations will be more precise, and LETKF assimilation will have seven steps.

2.2. Data assimilation in the WRF model

The data assimilation system in WRF was originally based on the MM5 3DVAR system and was named WRF3DVAR. Then a 4DVAR assimilation scheme was included and the name was changed to WRFVAR. In 2008 after the release of the hybrid variational/ensemble method, this component was renamed WRFDA, and is developed to this day.

2.2.1. Module đồng hóa WRFDA

WRFDA is used to feed observations into interpolation analyzes generated by the preprocessing system (WPS), which can also be used to update WRF initial conditions when running in cycle mode (two-way). WRFDA is based on an incremental variable data assimilation technique that supports both 3DVAR and 4DVAR methods. WRFDA also has mixed data assimilation capabilities that combine the benefits of the variational approach with the flow-dependent, statistical error information provided by aggregated forecasts.

The WRF-3DVAR assimilation method aims to generate an optimal estimate of the actual atmospheric state at the time of analysis by minimizing the cost function specified by the following function:

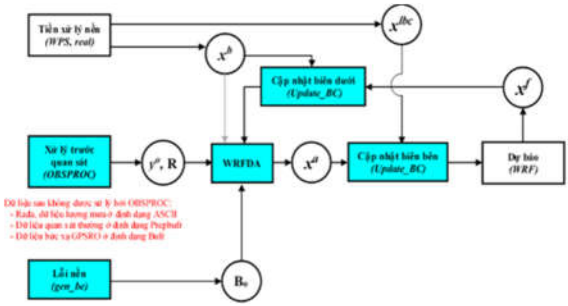

Figure 2.2. Relationship between WRFDA components and WRF system components

$$J(x) = J_b(x) + J_0(x) = \frac{1}{2}(x - x^b)^T B^{-1}(x - x^b) + \frac{1}{2}(y - y^0)^T (E + F)^{-1}(y - y^0)$$

(2.26)

3DVAR variable data assimilation is an iterative solution of equation (2.26) to find the analyzed state x such that *J(x)* is minimal. 4DVAR assimilation has a number of advantages over the 3DVAR method, allowing observations to be assimilated at the time of their observations or over a specific time period defining sub-predictive covariates depending on the traffic and most likely using the predictive model as a constraint leads to an improved analytic estimate. The algorithm function of the 4DVAR method has the form:

$$J(x_0) = \frac{1}{2}(x_0 - x_0^b)^T B^{-1}(x_0 - x_0^b) + \frac{1}{2}\sum_{i=1}^{n}\left[H_i(M_i(x_0)) - y_i\right]^T R_i^{-1}\left[H_i(M_i(x_0)) - y_i\right]$$

(2.28)

M_i is the predictive model and H_i is the observation operator over the estimated time divided by the fit interval *i*. In the problem of assimilating AOD data from MODIS satellite, the AOD value is the H_i. **B** is the baseline field covariance error matrix, which is a meteorological estimate, the vector background field x^b is the short-term forecast generated by a previous analysis. x^i is the vector representing the discontinuity analysis after the ***ith*** outer loop with $i = 1, \ldots, n$ where *n* is the number of iterations. x^n value vector obtained after the last (n*th*) outer loop. The inner loop optimization starts from a predictive state x^{n-1} which is the state of the analysis from the nearest outer loop. In the first outer loop, the background field x^b is usually taken as the first x^0 predictor state.

All observations within ±1 hour (for 3DVAR) and ±3 hours (for 4DVAR) will be processed, in which the observation data between 01 hour before and after the time of analysis (or 03 hours) are considered as observations during the analysis hour. This is illustrated in *Figure 2.4* and *Figure 2.5*.

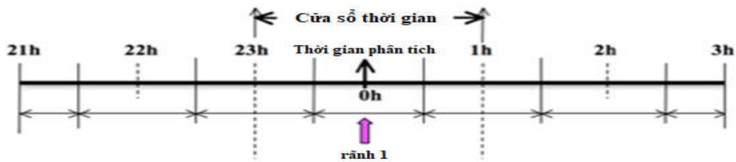

Figure 2.4: OBSPROC - 3DVAR Analysis Time Window

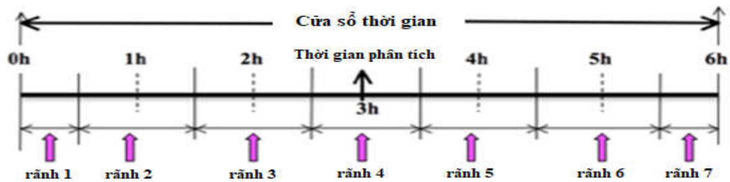

Figure 2.5: OBSPROC - 4DVAR Analysis Time Window

The AOD data from the MODIS satellite, after being processed in the above step, was assimilated into the WRF-CMAQ model system by the WRFDA module with the 4D-Var assimilation method. The process of assimilation of AOD data is carried out in step (4) in "Schematic of research steps and experimental simulation - *Figure 3.1*".

2.2.2. The Community Multiscale Air Quality Model CMAQ

Data after assimilation is converted into the CMAQ model through a meteorology - chemistry surface processor (Meteorology - Chemistry Interface Processor - MCIP).

In the CMAQ model system, data is processed through the following steps:

- Calculation of the photolysis rate used by the photochemical mechanism;
- ICON (Initial Conditions Processor): interpolate data according to the horizontal and vertical structure of the simulation domain.
- Boundary Conditions Processor (BCON): chemical conditions along the lateral boundaries of the simulation domain. BCON will generate an output file with chemical concentrations for all grid cells along the horizontal boundary of the modeling domain.
- CCTM chemical transport model (**C**MAQ Chemistry-**T**ransport **M**odel): input to CCTM includes data processed in the above steps and global emissions data determined by the chemical mechanism to calculate the chemical changes of the chemical composition on the basis of the variation of the meteorological factors, the result is the concentration of the chemical components (including PM particulate matter) expressed according to the meteorological factors on each time interval specified at the time step of the simulation.

Finally visualize the results of the simulation process in the WRF-CMAQ model system by Post-Processing modules.

Sub-conclusion of chapter 2

Data assimilation method was used for the purpose of combining observed data and model results to create the best quality input data set, improving the model's forecast results. Data assimilation technique using combinatorial Kalman filter uses filter parallelization very efficiently by dividing independent work pieces among different compute cores, which increase efficiency. The calculation has been applied in the WRF model system with the WRFDA data assimilation module.

The 4DVAR assimilation method in WRFDA based on the incremental variability data assimilation technique that allows observations to be assimilated at the time of observation or over a specific time period will be used in this study to observations into the interpolation analyzes generated by the Preprocessing System (WPS).

The MODIS 3 x 3 (km) resolution Aerosol product is assimilated through the WRFDA module after being divided into layers according to the altitude of the meteorological layer. It also solves the problem of the amplitude of variation of the data accordingly for each locality, filter observations, remove observations outside of simulation space and simulation time domain, assign quality flags to each observation, sort and merge duplicate data according to time and space, check the longitudinal consistency conditions of the simulation domain.

CHAPTER 3. RESULTS OF ASSIMILATED SATELLITE DATA FOR WRF-CMAQ MODEL SYSTEM

3.1. Determine the parameters for the experimental process

- Data source used:

The data used for the simulation include the following sources:

+ Meteorological data: Global coverage data from the US National Oceanic and Atmospheric Administration (NOAA).

+ Satellite data: AOD product is extracted from 3 km MODIS Aqua/Terra Aerosol 5-Min L2 Swath data via https://ladsweb.modaps.eosdis.nasa.gov/.

+ Global emissions data: Global emissions data are available on the CMAQ website.

- *Simulation steps:* Assimilation of satellite data for the WRF-CMAQ model system needs to be carried out through many different steps. Each implementation step requires many different stages, techniques, data, analysis and evaluation. The dissertation has performed experimental simulation steps as shown in *Figure 3.1*.

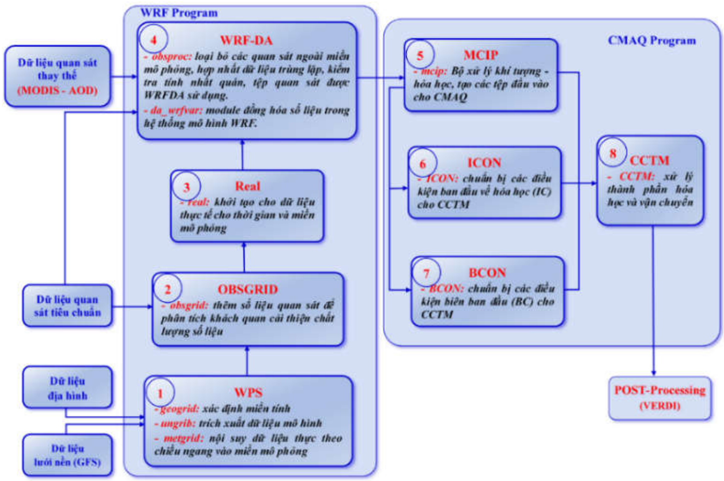

Figure 3.1. Flowchart of research steps and experimental simulation

- *Simulation time:* The time for the simulation is the period of time with stable weather conditions, AOD data from satellites is not affected by weather:

+ Rainy season:

From 00:00 16/01/2015 to 23:00 22/01/2015;

From 00:00 04/02/2015 to 23:00 11/02/2015;

From 00:00 05/02/2019 to 23:00 11/02/2019.

+ Dry season:

From 00:00 04/07/2017 to 23:00 on 10/07/2017;

From 00:00 06/09/2017 to 23:00 on 12/09/2017;

From 00:00 06/09/2019 to 23:00 12/09/2019.

- *Simulation experimental area:* The simulation area is Hanoi and its vicinity (*Figure 3.2*).

Figure 3.2: Domain (red) and nest (yellow) in Hanoi area

3.2. Simulation results

3.2.1. Simulation results in the rainy season

The flowing are the experimental simulation results during the rainy season including:

- From 00:00 16/01/2015 to 23:00 22/01/2015;
- From 00:00 04/02/2015 to 23:00 11/02/2015;
- From 00:00 05/02/2019 to 23:00 11/02/2019.

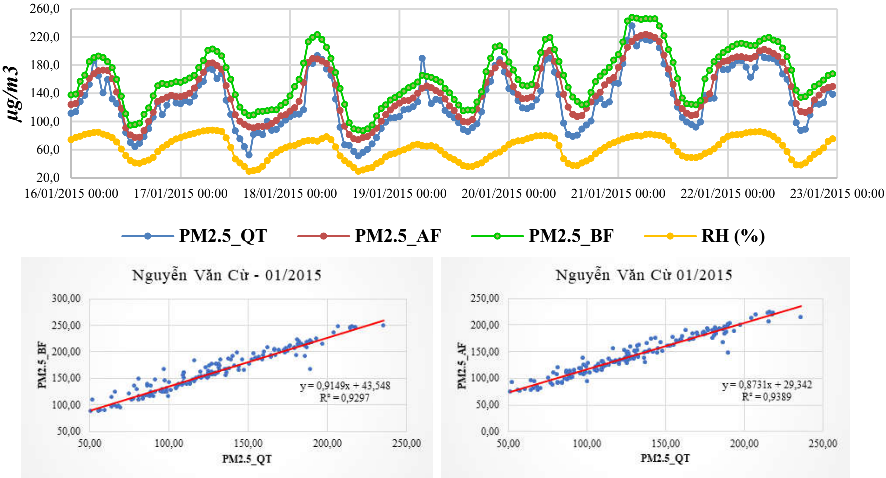

Figure 3.12: Correlation, regression results of $PM_{2.5}$ concentration from monitoring station and CMAQ model before and after assimilation from 00:00 16/01/2015 to 23:00 22/01/2015 at Nguyen Van Cu station

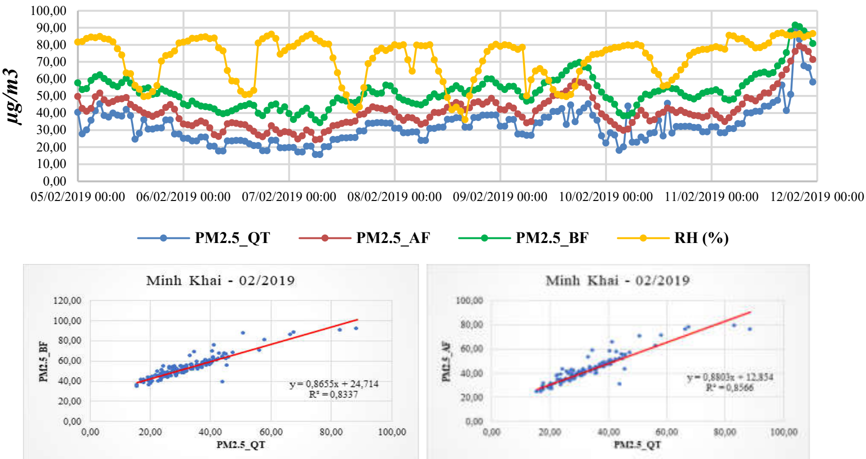

Figure 3.15: Correlation, regression results of $PM_{2.5}$ concentration from monitoring station and CMAQ model before and after assimilation from 00:00 05/02/2019 to 23:00 11/02/2019 at Minh Khai station

Regression results have shown that the correlation between $PM_{2.5}$ concentration before and after assimilation with improved $PM_{2.5}$ concentration (R^2 from 0.929 to 0.938 at Nguyen Van Cu station in January 2015 - *Fig. 3.12*; and from 0.833 to 0.857 at Minh Khai station February 2019 - *Fig. 3.15*).

3.2.2. Simulation results in dry season

Following are the experimental simulation results during the dry season including:

- From 00:00 04/07/2017 to 23:00 on 10/07/2017;
- From 00:00 06/09/2017 to 23:00 12/09/2017;
- From 00:00 06/09/2019 to 23:00 12/09/2019.

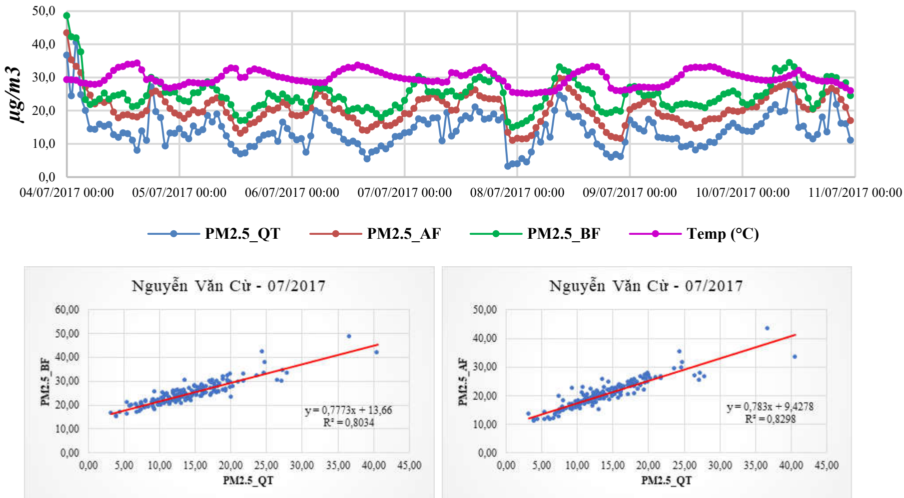

Figure 3.16: Correlation, regression results of $PM_{2.5}$ concentration from monitoring station and CMAQ model before and after assimilation from 00:00 04/07/2017 to 23:00 10/07/2017 at Nguyen Van Cu station

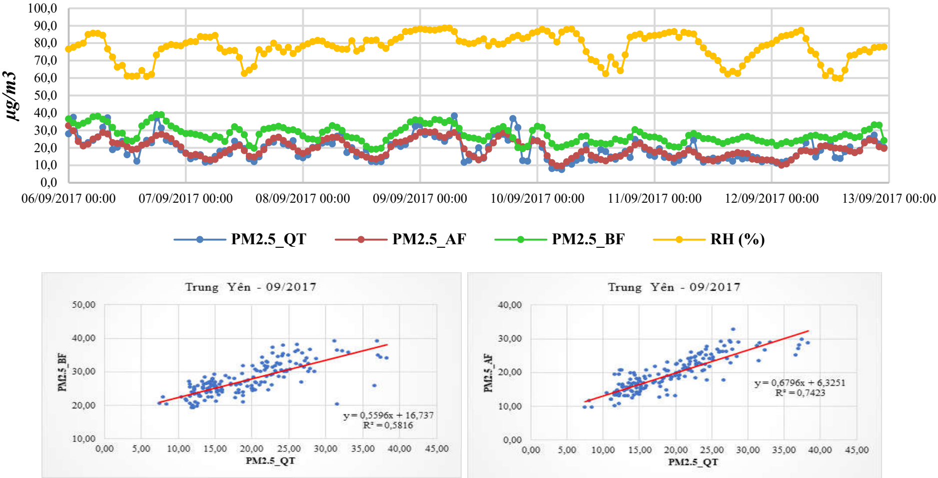

Figure 3.19: Correlation, regression results of $PM_{2.5}$ concentration from monitoring station and CMAQ model before and after assimilation from 00:00 06/09/2017 to 23:00 12/09/2017 at Trung Yen station

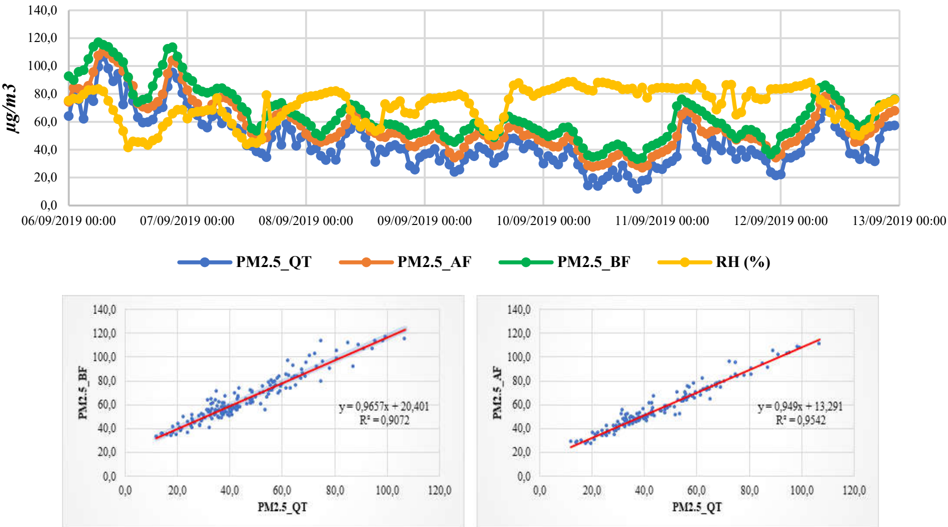

Figure 3.23: Correlation, regression results of $PM_{2.5}$ concentration from monitoring station and CMAQ model before and after assimilation from 00:00 06/09/2019 to 23:00 12/09/2019 at Minh Khai station

The regression results show that the correlation between $PM_{2.5}$ concentration before and after assimilation with improved $PM_{2.5}$ concentration (R^2 from 0.803 to 0.829 at Nguyen Van Cu station in July 2017 - *Fig. 3.16* and from 0.582 to 0.742 at Trung Yen station in September 2017 - *Fig. 3.19*).

The estimated $PM_{2.5}$ concentration from the model after assimilation also shows that the correlation has improved significantly, R^2 increased from 0.907 to 0.954 at Minh Khai station in September 2019 - *Fig. 3.23*.

The simulation results clearly show that the AOD satellite data has been assimilated in the simulation results and has a positive impact on the estimated $PM_{2.5}$ concentration in particular and the chemical concentration fields in general.

Sub-conclusion of chapter 3

The simulation of the experiment was carried out on the entire Hanoi area and its vicinity, divided into 06 simulations with a duration of 07 days each, divided by dry and rainy seasons for the years 2015 (02 dry season simulations), 2017 (02 simulation of rainy season) and 2019 (01 rainy season simulation and 01 dry season simulation).

The process of assimilating AOD optical depth data at 550 nm extracted from MODIS satellite data of experimental simulations is performed through the WRFDA module in the WRF model system that has built the first set of input data for the CMAQ air quality model with high quality.

With the data source after assimilation by the WRFDA module, the CMAQ air quality model has simulated $PM_{2.5}$ concentration for the

Hanoi area with the simulation time period. The $PM_{2.5}$ concentration from the results of the CMAQ model has a significantly improved correlation with the ground observation data compared to the data without assimilation. For particular: the R^2 regression results increase at least from 0.915 increased to 0.917 for the simulation from 06/09/2017 to 12/09/2017 at Nguyen Van Cu station and the maximum from 0.581 to 0.742 for the simulation time from 06/09/2017 to 12/09/2017 at the Trung Yen station; R2 increased from 0.929 to 0.939 for the simulation from 16/01/2015 to 22/01/2015 at Nguyen Van Cu station; from 0.803 to 0.830 for the simulation from 04/07/2017 to 10/07/2017 at Nguyen Van Cu station; from 0.368 to 0.431 for the simulation from 04/07/2017 to 10/07/2017 at Trung Yen station. For the simulation periods 2017 and 2019 at Trung Yen station and Minh Khai station, the regression results both increased significantly.

Regression results of $PM_{2.5}$ concentration with and without assimilation of AOD data with $PM_{2.5}$ concentration at ground monitoring stations during simulation time periods both increased, showing a good correlation with meteorological factors as well as atmospheric chemistry. Experimental results suggest that assimilation has the potential of providing significant improvements to atmospheric chemistry prediction.

However, a major limitation of satellite data assimilation is the limited availability of data, especially in real time (or near real time). Assimilation of surface data and satellite data has proven useful. Satellite data is valuable due to the contemporaneous coverage it can provide.

CONCLUSIONS AND RECOMMENDATIONS

Conclusions

The application of the combined Kalman filter in the WRFDA data assimilation system of the WRF weather forecasting and research model has improved the estimation accuracy, significantly improved the initial state, and provided the input data for CMAQ air quality model, improving the results of air quality forecast for Hanoi area.

AOD data processing software to remove raw, substandard, inconsistent data, flags based on data reliability has been built, the application has clear results in a test simulation for the Hanoi area.

The benefit of assimilated MODIS satellite data applied in the WRF-CMAQ modeling system improves the accuracy of model inputs (IC, BC, and emissions) and forecasts. The AOD data from the MODIS satellite that was assimilated into the CCTM has improved performance on surface air pollution concentrations and better reanalyzed air quality.

The assimilation of aerosol data also known as optical depth extracted from the 3 km MODIS Terra/Aqua Aerosol 5-Min L2 Swath satellite data significantly improves the $PM_{2.5}$ concentration estimation results on the WRF-CMAQ model system. Model performance was improved with this metric assimilation.

The research has built a process of assimilating AOD satellite data, providing input data sources for the CMAQ air quality model for forecasting air quality in accordance with Vietnam conditions when the density of direct monitoring stations is sparse, especially important for areas where monitoring stations have not been installed.

In general, the assimilation of satellite data for air quality modeling provides significant benefits when the traditional monitoring data are only in point form (although there may be many monitoring stations) and on the ground surface is not enough to meet the requirements of the air quality assessment. Assimilation of satellite data for air quality modeling opens up a new way to provide air quality assessment solutions.

Recommendations

Based on the research results, a number of recommendations are made with the aim of creating tools for monitoring the air environment, specifically as follows:

- Develop emission data sets for Vietnam and surrounding areas to provide input data for the CMAQ model, improve the quality of simulation results, and forecast air quality for Vietnam.
- Research and application of satellite images with higher resolution (both in space and time) to take advantage of multi-spectral, multi-time and instantaneous remote sensing on a large scale and simultaneously for the WRF-CMAQ model system. In order to enhance the quality and performance of the model in assessing air quality for managers to plan policies and improve air quality for the community.
- The results of this research need to be developed to be applied in air quality assessment to serve the assessment and forecasting of trends in order to adjust human activities to contribute to improving air quality as well as community health.
- In addition to the WRF-CMAQ model system, currently the WRF-Chem model has been developed by the user community and growing, therefore the research of satellite data assimilation for this model is also the goals that needs to be studied and put into practice in reality.

LIST OF PUBLISHED WORKS RELATED TO THE DISSERTATION

1. 1.Tran Dang Hung, Doan Ha Phong, Hoang Thanh Tung, Nguyen Ngoc Anh, Le Phuong Ha, Nguyen Thi Minh Hang, Nguyen Ngoc Kim Phuong, **Nguyen Hai Dong** (2017), "Application of GIS technology and monitoring satellites of changes in $PM_{2.5}$ concentration in northern Vietnam (2000 - 2005 - 2010)", *Collection of National Scientific Conference on Meteorology, Hydrology, Environment and Climate Change*, Institute of Meteorology, Hydrology and Climate Change, p. 476-482.
2. **Nguyen Hai Dong**, Doan Ha Phong (2020), "Empirical relationship between $PM_{2.5}$ and aerosol optical depth (AOD) in Hanoi urban city", *Vietnam Journal of Hydrometeorology*, No. 718 October, 2020. ISSN 2525 - 2208.
3. **Nguyen Hai Dong**, Doan Ha Phong, Le Ngoc Cau (2020), "Application of 4DVAR method to assimilate AOD data from MODIS satellite to forecast $PM_{2.5}$ concentration in Hanoi area", *Journal of Climate Change Science*, No. 16, Q4 2020. ISSN 2525 - 2495.