Estimations of $PM_{2.5}$ concentrations based on the geographically weighted regression from Himawari-8 AOD

Wende Zhu[1,2], Qiushuang Zhang[2], Kun Cai[2,a], Lei Wang[2] and Shenshen Li[1]

[1] State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100020, China

[2] Computer and Information Engineering College, Henan University, Kaifeng 475004, China

[a] Kun Cai: henu_caikun@126.com

Abstract. Polar-orbit satellites have the limitation of observed frequency, so the method of $PM_{2.5}$ concentration estimation based on MODIS AOD is difficult to capture the variation in trend and distribution characteristics of $PM_{2.5}$ in a day. Himawari-8 geostationary satellite with the advantage of high temporal resolution can obtain the hourly aerosol optical depth(AOD).This paper selected AOD from 2km Himawari-8 Standard Data to estimate the $PM_{2.5}$ concentration based on Geographically Weighted Regression(GWR)model, and used the ground-level $PM_{2.5}$ concentrations as training samples of the model. In addition to AOD, the model also took meteorological data such as height of the planetary boundary layer (HPBL) and relative humidity (RH) as parameters to participate in model training to improve the reliability of the results. At the end of the paper, it utilized the model to estimate $PM_{2.5}$ concentration over eastern China in July 2018 and executed the validation with ground-level $PM_{2.5}$ concentrations. The result indicated there is a basic consistency in the change trend between them (R=0.74, RMSE=10.07$\mu g/m^3$).

1.Introduce

With the acceleration of industrialization and urbanization, $PM_{2.5}$ has become one of the harmful pollutants affecting China's air quality which is a particulate matter with an aerodynamic diameter less than or equal to 2.5μm. Because $PM_{2.5}$ has a small particle size and contains a large number of toxic substances, most of the $PM_{2.5}$ entering the human body is deposited in the bronchi and alveoli, which may directly endanger human health[1-2].

In the field of atmospheric remote sensing. AOD as the precursor of $PM_{2.5}$, is usually obtained from the satellite products and used to estimate $PM_{2.5}$ concentration based on AOD-$PM_{2.5}$ correlation model .Wang et al. compared the ground-level $PM_{2.5}$ concentration of seven sites in Jefferson county with the corresponding MODIS AOD, it is found that the correlation coefficient between them reaches 0.7, which further proved that it is feasible to establish the correlation between $PM_{2.5}$ and AOD[3]. This kind of linear regression model is relatively simple and intuitive. However, as the interference of the atmospheric vertical distribution on the model is not taken into account, so once the study area is expanded, such models have great uncertainty in prediction accuracy. Therefore some researchers have introduced such meteorological factors as HPBL, RH, temperature to be auxiliary variables combined

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

the AOD to construct a multivariate linear regression model for estimate $PM_{2.5}$ concentration. Liu et al. adopted GASP AOD, meteorological factors and combined other factors such as population density and land-use type, used the method of regression analysis twice to obtain the $PM_{2.5}$ distribution in the eastern United States. This method can better predict the spatial and temporal changes of $PM_{2.5}$, and the inter-model is applied to the estimation of $PM_{2.5}$ for the first time[4].

In summary, although researchers have conducted AOD-PM2.5 research based on MODIS AOD normally due to the very high spatial resolution, it also has a defect with relatively low temporal resolution, for instance, only 1-2 coverage observations per day can be achieved for the same area[5]. As a comparison, the Himawari-8 geogeostationary satellite can provide observe at a frequency of every 10 minutes[6].Therefore, in order to satisfy the demand for obtaining $PM_{2.5}$ concentration data with high temporal resolution, this paper was based on the Himawari-8 AOD and utilized GWR to estimate $PM_{2.5}$ concentrations.

2.Material and Methods

2.1 Data Source

2.1.1 Himawari-8 AOD

Himawari-8 geogeostationary satellite has the characteristics of high temporal resolution and can be used to study the daily or even hourly changes of AOD. With the scanning of Full Disk, the observation period of a satellite image in a fully covered area is 10min, and 144 observations can be made in a day which embody the high flexibility in time. On account of the Himawari-8 official just provides 5km AOD product, this paper based on the dark pixel algorithm to establish an accurate surface reflectance relationship to retrieve the 2km Himawari-8 Standard Data released by Japan Aerospace Exploration Agency(JAXA)[7].

This paper chose the Aerosol Robotic Network(AERONET)AOD(Level 2.0) to validate the above AOD results. Since ground-level $PM_{2.5}$ concentration the training samples and validation criteria, is hourly average, Therefore this paper used the above hourly AOD to be the input parameters of the AOD-$PM_{2.5}$ GWR model. By comparing the hourly valid values of AOD from 9:00AM to 17:00PM with the corresponding hourly mean of AEROENT whose time range was July 2018. The Table 1 showed that the retrieval results were relatively stable and had little fluctuation. In conclusion, all these statistical parameters proved that the retrieval quality was reliable.

Table 1. Statistics for Himawari-8 AOD and AERONET AOD. The absolute error(Diff), Relative Error(RE), expected error(EE), double expected error (2EE) are included in the table.

Dataset	Mean	Diff	RE(%)	EE(%)	2EE(%)
AERONET	0.501	0.04	8.28	67.6	87.4
Himawari-8	0.536				

2.1.2 Meteorological and ground-based data

HPBL and RH exert a critical effects on the diffusion and concentration of atmospheric pollutants, which are important meteorological factors that determine the spatial and temporal distribution of $PM_{2.5}$ concentration. In this paper, adopted the reanalysis data of assimilation system which developed by the National Weather Service's National Centers for Environmental Prediction(NCEP).The obtained HPBL and RH from this system with a temporal resolution of 3h and a spatial resolution of 0.5 degree can well match himawari-8 AOD to be the meteorological parameters of AOD-$PM_{2.5}$ GWR model.

Since January 2013, the air-quality monitoring stations established by China Environmental Monitoring Center across the country has started to release ground-based pollutant data one after the other. So far that China has a total of 1,809 monitoring stations. In order to effectively build the relationship between AOD and $PM_{2.5}$, this paper selected the eastern China with relatively dense sites as the research object. In this paper, used the hourly ground-level $PM_{2.5}$ concentrations to be the training sample for the GWR model. In the discussion part, it would be used as evaluation standard to compare with the $PM_{2.5}$ concentrations which estimated by the GWR model.

2.2 Methods

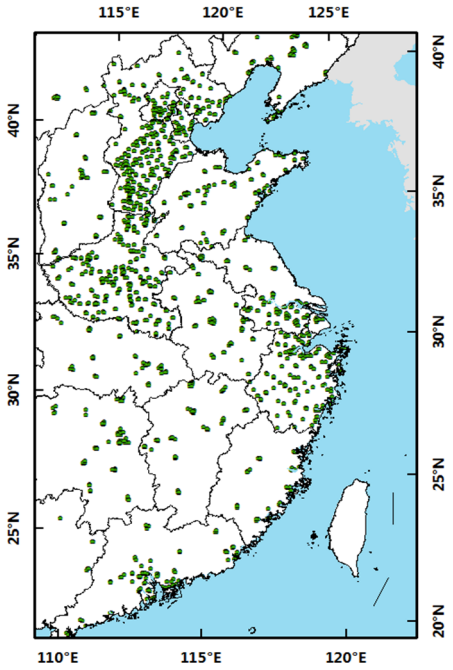


Figure 1. Distributionof air-quality monitoring stations in eastern China

2.2.1 Spatial and temporal matching of data

In order to reduce the impact of data noise and spatial instability, such AOD, HPBL, RH, and ground-level $PM_{2.5}$ concentration as input parameters should be effectively matched in time and space before building the GWR model. In time scale, set the AOD' time as the benchmark, chose the $PM_{2.5}$ concentration average of three hours which include the corresponding time , the previous hour and the later hour. Meteorological data is selected at the closest time .On the spatial scale, Set the location of ground-based monitoring station as the centre, chose the AOD' average of 25 pixels around the centre. The pixel of meteorological data was selected where is nearest to the centre.

2.2.2 GWR

The general linear regression model used to fit the relationship between two variables tend not to reflect the altered characteristics of spatial data veritably. To solve this problem, Brunsdonet et al. proposed the concept of Geographical Weighted Regression model [8-9]. This is a kernel regression technology that can capture the spatial nonstationarity by calibrating multiple regression models and reflect the changing characteristics of spatial relations between variables within the region.

This paper structured the Himawari-8 AOD –$PM_{2.5}$ GWR based on the MODIS AOD-$PM_{2.5}$ GWR

proposed by Chen et al. that utilize the prior relationship between AOD and PM25 concentration to deduced the general formula among AOD, HPBL, RH and PM25 concentrations[10-11].

$$\begin{aligned} &ln\, PM_{2.5}(u_i, n_i) \\ &= \beta_0 + \beta_1(u_i, n_i)\, ln\, AOD \\ &\quad + \beta_2(u_i, n_i)\, ln\, PBLH \\ &\quad + \beta_3(u_i, n_i)\, ln(1 - \frac{RH}{100}) \end{aligned} \qquad (1)$$

Assuming that there are i groups of ground-based monitoring stations in different positions, where β was a regression coefficient between each input parameter and $PM_{2.5}$ concentration in the observation point(u_i , n_i).In order to detect spatial changes, introduced the weighted least square method to estimate the value of β. Used the matrix to describe the formula (1):

$$\beta = \ (X^T W X)^{-1} X^T W Y \qquad (2)$$

Where X was a four-dimensional aggregate included the value of $\ln AOD$, $\ln PBLH$, $\ln(1 - \frac{RH}{100})$ and β_0 in each ground monitoring station. In the case of empirical relations, usually set β_0 equal to 1, Y was an aggregate included the value of $\ln PM_{2.5}$, W was a spatial weight matrix. This model used the Gaussian Function to be kernel function to calculate the spatial weight matrix.it was defined as:

$$W_{ij} = exp\ (-(^{d_{ij}}/_{b})^2 \qquad (3)$$

Where d_{ij} was the distance between station i as a regression point and the other station j, and b was a nonnegative distance attenuation parameter that described the function relationship between weight and distance, called Band Width. Different Band Widths will generate different spatial weight matrixs. For a given d_{ij} . If b is larger, the weight of the observed value at j position will be smaller, on the contrary that weight will be larger. This paper adopted the method Cross-Validation[8-9] proposed by Brunsdonet et al. to choose the optimal Band Width.The method is described as :

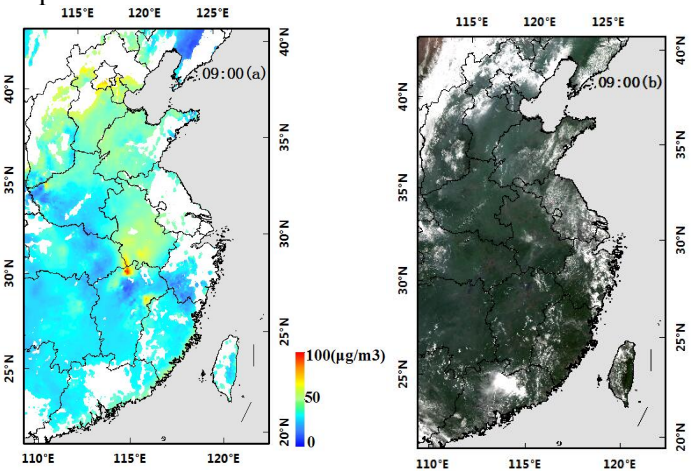
$$\mathrm{CV}_j = \sum_{i=1}^{n} [Y - Y']^2 \qquad (4)$$

$$CV_{opt} = \mathrm{MIN}(CV_1, CV_2, \ldots, CV_j) \qquad (5)$$

Where $Y'$was the $PM_{2.5}$ concentration fitting results .By means of taking the initial Band Width to the formula (3),and through the iterative calculation with (1) and (2) to obtain it.

As it is shown in the (5), the value of b for CV_{opt} is the optimal Band Width. Taking the filtered b to repeat the above calculation to achieve the regression coefficient in case of the optimal Band Width. So far the construction of GWR model has completed.

Finally, through the formula (1) with the above regression coefficient combining the distance between every valid AOD pixel and nearby ground monitoring station to estimate the $PM_{2.5}$ concentration results for all pixels where has an AOD valid value.



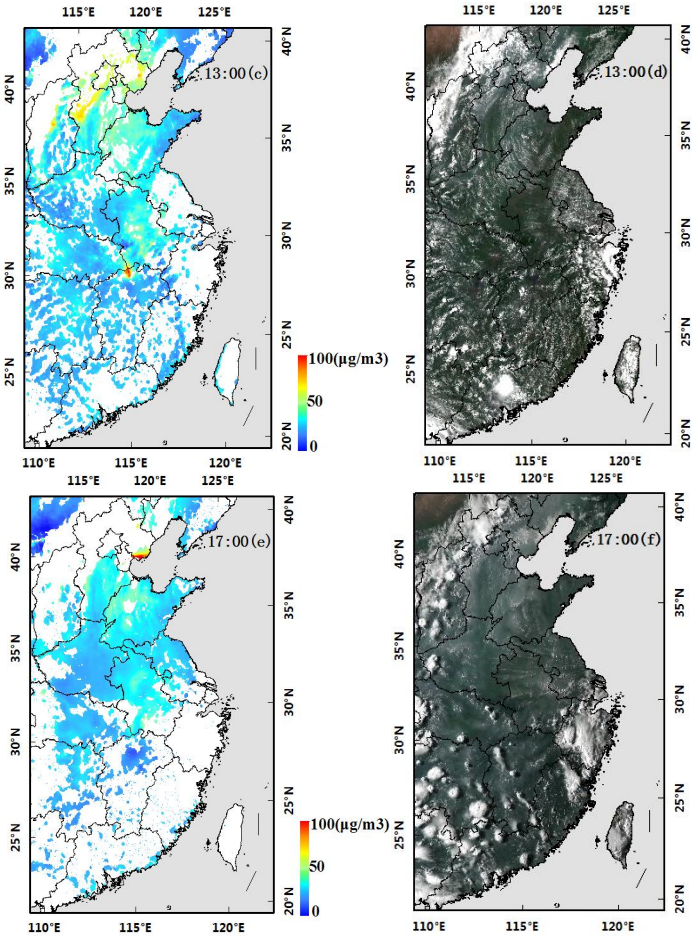


Figure 2. The $PM_{2.5}$ concentrations images and true-color images at 09:00LT, 13:00LT and 17:00LT.

3.Results and Discussion

In this section, this paper conducted the estimation of $PM_{2.5}$ concentration with Himawari-8 AOD based on the algorithm described in section 2.2.2, and a certain cases were be shown in the following part. Furthermore, the validation between ground-level and estimated $PM_{2.5}$ concentration was also executed in this section.

The hourly AOD from 9:00AM to 17:00PM in July 2018 matched the other parameters with the methods in section 2.2.1, to conduct the $PM_{2.5}$ estimation over eastern of China. Here are some samples. The left side of Figure 2 presented the spatial distribution of $PM_{2.5}$ concentration on 21 July 2018 at 09:00 Local Time(LT), 13:00LT and 17:00LT. The right is the corresponding true-color image at the same time. By contrast, the estimated $PM_{2.5}$ distribution images was close to the right true-color images and the coverage area of estimated results was influenced by the cloud obviously. In addition, at 09:00LT, there was most widespread $PM_{2.5}$ distribution and the overall concentration value was also relatively high and that gradually descended after noon.

Figure 3 presented the scatterplot of the comparison between ground-level and estimated $PM_{2.5}$ concentrations over eastern China in July 2018(09:00–17:00LT). On the time scale, set the estimated $PM_{2.5}$ concentrations to match the ground-level $PM_{2.5}$ concentrations hourly. In order to reduce the impact of abnormal values caused by satellite signal error, this paper chose the average of valid estimated $PM_{2.5}$ concentrations at a spacial range of 10km×10km (25pixels) around the ground-based monitoring stations. The matching results amounted to 168981 sample points.

Linear correlation analysis was performed for the fitting results, as shown in Figure 3(a). There was a close average between them. And the correlation coefficient reached 0.74, RMSE of 10.07($\mu g/m^3$). It can be seen that there was a moderate consistency and error control for demonstrating that by means of GWR model to construct the relationship between AOD and $PM_{2.5}$ is feasible. However the slope was just 0.57, intercept of 12.23, indicated that there are some cases where high values were underestimated or low values were overestimated apparently. And this phenomenon was more obvious in Figure 3(b) in the case of selected hours. The latter one may cause by cloud identification algorithm for AOD wasn't enough accurate at higher latitudes. Due to the high reflectivity of cloud pixels, if it wasn't eliminated before participating the retrieval algorithm, the AOD of corresponding pixel were be seriously overestimated which would cause $PM_{2.5}$ concentrations were overestimated. In the selected three moments of the day, Figure 3(b) shown the highest R(0.75) and RMSE(11.28 $\mu g/m^3$), as time goes on, those reached the lowest(R=0.70, RMSE=7.58 $\mu g/m^3$) at 17:00LT shown in the Figure 3(d). Meanwhile, there was the minimum number of samples(13780)at 17:00PM as the sun sets causing the satellite scanning area to shrink.

4.Conclusions

This paper was based on AOD retrieved from Himawari-8 Standard Data and combined GWR model to simulate $PM_{2.5}$ concentrations, presented and verified the results in July 2018 over eastern China .The conclusion was shown as below.

(1)The method of $PM_{2.5}$ estimation based on Himawari-8 AOD has exploited the advantages to the full of high temporal resolution of geogeostationary satellite. It can satisfy the need of real-time observation. It can be seen from the comparison of Figure 2 that the experimental results can better reflect the pollution degree and the characteristics of spatial distribution in the moment.

(2) In section 3. Through taking the ground-level results to verify estimated results, this paper can come to the conclusion that the latter one had moderate correlation with the former one. But the stability and precision of the algorithm still need to be improved. In terms of AOD, although the geostationary satellite has the advantage of high temporal resolution, it is still difficult to reach the accuracy level of MODIS, among which the official cloud products are not accurate enough in the high latitude region may cause the phenomenon that the low values were overestimated.

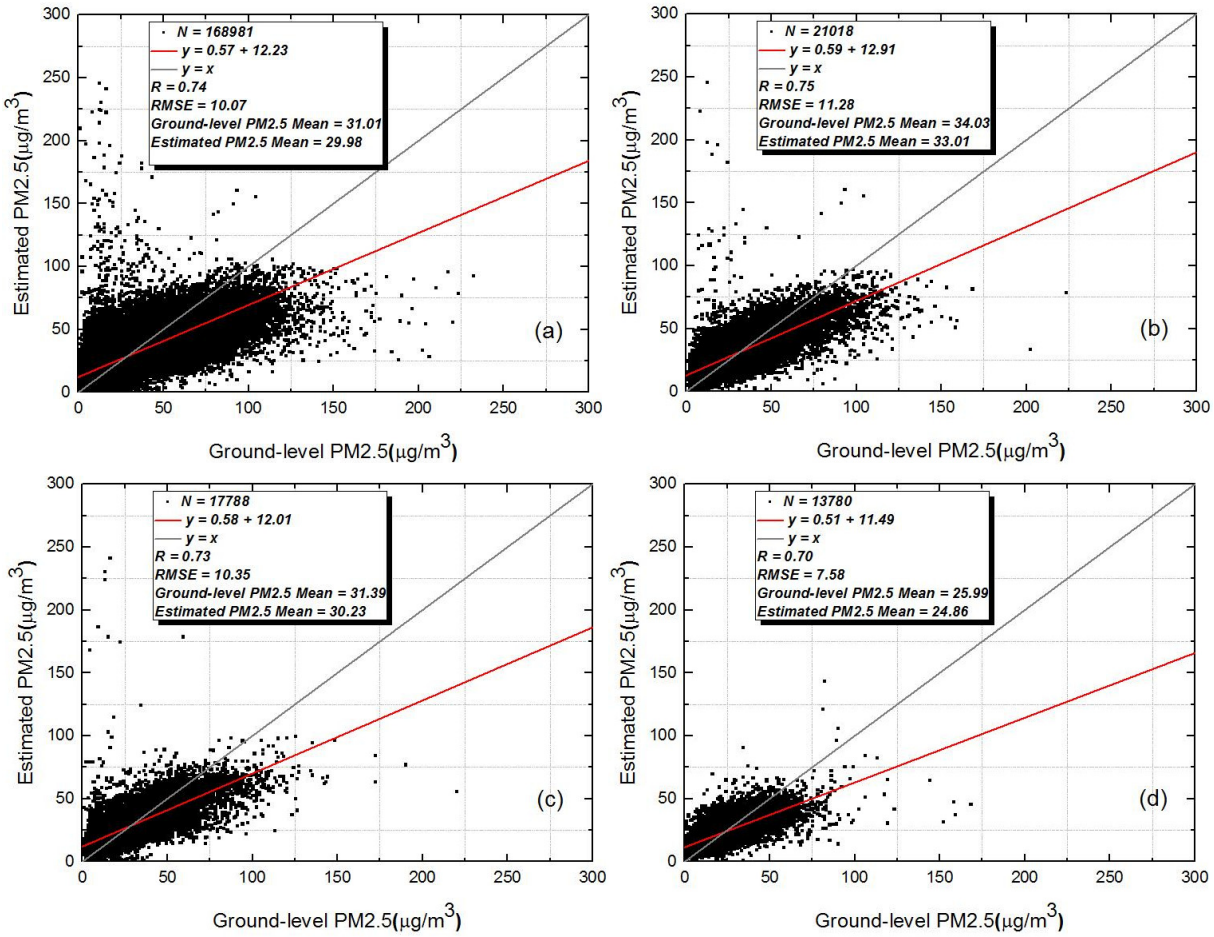


Figure 3. The scatterplot of estimated $PM_{2.5}$ concentration by comparing ground-level $PM_{2.5}$ concentrations from all available data(a) and in different hours 09:00LT(b), 13:00LT(c), 17:00LT(d).

Acknowledgements

Thanks to the China National Environmental Monitoring Centre for providing $PM_{2.5}$ data. This work was supported by the National Key R&D Program of China (2016YFC0201507), and the National Natural Science Foundation of China (Grant No. 41471367). The authors acknowledge the Himawari-8 data support from JAXA.

References

[1] Yang H, Zou X, WANG H, et al. Study progress on $PM_{2.5}$ in atmospheric environment［J］. Journal of Meteorology and Environment, (2012), **28(3)**: 77－82.

[2] Takeshi F, Hayashi S,Hogg JC, et a1. Particulate matter induces cytokine expression in human bronchial epithelial cells[J]. American Journal of Respiratory Cell and Molecular Biology, (2001), **25(3)**：265-271.

[3] Wang J, Christopher S A. Intercomparison between satellite-derived aerosol optical thickness and $PM_{2.5}$ mass: Implications for air quality studies[J]. Geophysical Research Letters, (2003), **30(21)**: 267-283.

[4] Liu Y, Paciorek C. J, Koutrakis P K. Estimating Regional Spatial andTemporal Variability of PM2．5 Concentrations Using Satellite Data,Meteorology,and Land Use information[J]. Health Perspect, (2009), **117(6)**: 886-892.

[5] King M D, Menzel W P, Kaufman Y J, et al. Cloud and aerosol properties, precipitable water, and profiles of temperature and water vapor from MODIS[J]. IEEE Transactions on Geoscience & Remote Sensing, (2003), **41(2)**:442-458.
[6] Bessho K, Date K, Hayashi M, et al. An Introduction to Himawari-8/9— Japan's New-Generation Geostationary Meteorological Satellites[J]. Journal of the Meteorological Society of Japan. (2016), **94 (2)**: 151-183.
[7] Otemachi, Chiyoda K. Himawari-8/9 Himawari Standard Data User's Guide Version 1.2[R]. Tokyo：Japan Meteorological Agency. (2015).
[8] BRUNSDON C, FOTH ERINGH AM AS, CHARLTON M .Geographically weighted regressi on :A method f or exploring spatial[J]. Geographical Analysis, (1996), **28(4)**: 281-298.
[9] BRUNSDON C, FOTH ERINGH AM AS, CHA RLTON M .Spatial Nonstationarity and Autoregressive Model[J]. Environment and planning A, (1998), **30**: 957-973.
[10] Chen H, Li Q, Zhang YH, et al. Estimations of PM2．5 concentrations based on the method of geographically weighted regression[J]. Acta Scientiae Circumstantiae, (2016), **36(6)**: 2142-2151.
[11] JIA S, Su L, TAO J, et al. A study of multiple regression method for estimating concentration of fine particulate matter using satellite remote sensing[J]. China Environmental Science, (2014), **34(3)**: 565~573.