A Comparison of Spatial Interpolation Methods for predicting concentrations of Particle Pollution (PM_{10})

Phatarapon Vorapracha*, Pongtep Phonprasert, Suparada Khanaruksombat and Nuchanaporn Pijarn

***Abstract*—**The aim of this experiment was to compare the performance of three interpolation method as Inverse Distance Weighting (IDW) method, Ordinary kriging (OK) method and Universal Kriging (UK) method for predicted the air pollution condition (PM_{10}) in the center of Thailand. The interpolated with each interpolation method for 20 air monitor station for 24 mouths (January 2013 - December 2014) show that IDW method is the best interpolation method to predicted the air pollution condition with the Root Means Square Error (RMSE) as 0.683-0.703.

***Keywords*—**Interpolation, Inverse Distance Weighting, Ordinary kriging, Universal Kriging

I. INTRODUCTION

NOWADAYS, air pollution is the major problem and more deteriorating. The increasing of world population and the changes of people lifestyle cause increasing power consumption e.g. industrial, transportation and electricity; which are all resulting in air pollution. The amount of air pollutant that is emitted appears in the form of small droplets called total suspended particles (TSP). Particulate matter (PM_{10}) is the main component in TSP problem. If air pollution is responsible for the observed increased mortality, one would expect to see it impact on human respiratory systems especially the high risk groups, including children, elder and people with congenital respiratory diseases.

The central of Thailand is the most densely populated area. The topography mainly consists of a plain which is caused by the river sediment deposition for millions of years. Thailand central plains are ranged from the south of Uttaradit province down to the Gulf of Thailand. It is a wide-ranging plain than the other regions of the country. However, there are mountains on some areas of the plains as the mountain in Nakhon Sawan province and the western side of Phitsanulok province. From the geological evidence, it can assume that these mountains used to be the islands from the world flood in Medieval Warm Period.

Pollution Control Department (PCD) [1] has been measuring the amount of particle smaller than PM_{10} but it does not cover all area because the limited of measurement tools and expansive survey cost. This research can predict these partial beside the area measurement by comparing between Inverse Distance Weighting [2], Ordinary Kriging and Universal Kriging method [3] of central plain of Thailand to be used in planning and prevented the air pollution control and solve air pollution problem in the future.

Phatarapon Vorapracha is with Division of Information Technology, Faculty of Science and Technology, University of BangkokThonburi, 10170, Thailand (e-mail: phatarapon@yahoo.com)

PongtepPhonprasert is with Division of Environment Technology, Faculty of Science and Technology, University of BangkokThonburi, 10170, Thailand (e-mail: pongtep_ph@outlook.com)

SuparadaKhanaruksombat is with Division of Environment Technology, Faculty of Science and Technology, University of BangkokThonburi, 10170, Thailand (e-mail: supee.khana@gmail.com)

NuchanapornPijarn is with Division of Environment Technology, Faculty of Science and Technology, University of BangkokThonburi, 10170, Thailand (e-mail: nuchtennis@gmail.com)

II. MATERIAL AND METHOD

A. Study area and data

The study area of this research was carried out in the Central region of Thailand. The main data source of this paper comes from Thai Meteorological Department includes monthly mean surface temperature data at 30 stations distributed in this region and spatial distribution classify by province during the period 2013 to 2014. This observed data set consist station name, latitude, longitude, and temperature values. For the purpose, we used inverse distance weighting (IDW), Ordinary Kriging (OK) and Universal kriging (UK) interpolated Particle Pollution data for estimated unknown points in this region.

B. Inverse Distance Weighting

Inverse Distance Weighting (IDW) method has used the concept of Tobler's first law to explain base on "It was defined as everything is related to everything else, but near things are more related than distant things." [4]. IDW method was an exact local non-geostatistics spatial interpolation method. It estimates values for the unknown values of the summation linear of points with known values. Then, we can weight by the inverse of the distance between points with known values for the unknown values. IDW method can be calculated.

$$u(x) = \sum_{i=0}^{N} \frac{w_i(x)u_i}{\sum_{j=0}^{n} w_j(x)}$$

$$w_i(x) = \frac{1}{d(x, x_i)^p}$$

Where *u(x)* are estimated values of unknown points, *wi(x)* are IDW weighting functions, *ui*are values of known points, *x* are interpolating unknown points, *xi* are interpolating knownpoints, *d* are distance from the known points *xi* to unknownpoints *x*, *N* are the number of known points and *p*are powerparameters used in interpolation. By, *p* is the main factoraffecting the estimation of IDW and *p* values are will focus onpoints near *p* values less to focus on points beyond.

C. Ordinary Kringing

Ordinary Kriging is a spatial estimation method where the error variance is minimized. This error variance is called the kriging variance. It is based on the configuration of the data and on the variogram, hence is homoscedastic. It is not dependent on the data used to make the estimate. Recently, Yamamoto derived error variance for ordinary kriging that is conditional to the data values. He referred to this variance as the Ordinary Interpolation variance. Yamamoto has shown that the ordinary interpolation variance is a better measure of accuracy of the kriging estimate. The ordinary kriging programs used for this study were modified to calculate the new error variance, named the Ordinary kriging interpolation variance (NKVAR) and output it along with the traditional kriging variance. The 95% error bound based on the new variance was reported also. It is believed that the new method used in this study to determine the interpolation variance is a better estimate of the error variance than the kriging variance. In particular, for skewed data, it is believed that the new variance is a much better estimate of the error variance.

Ordinary Kriging method was estimated unknown values from the known values. By, there is a half of the average variation of the point we had known the value (average semi-variance). It was a spatial relationship of the data is in the same line. There is average semi-variance weighting parameter to calculate interpolated value of unknown points from semi-variance function and basic equation for estimate z value of a point follow

$$\gamma(h) = \frac{1}{2n} \sum_{i=1}^{n} [z(x_i) - z(x_i + h))]^2$$

When γ (*h*) is average semi-variance value, *h* are distance between points, *n* are pair numbers of points refer from distance and *z* are values of known points.

$$Z_0 = \sum_{i=1}^{S} z_x w_x$$

When $Z_0$are estimate values, $Z_x$are values of known points, $W_x$are correlated weighting values between known points withunknown points and *s* are number of known points for estimation.

D. Universal Kriging

Universal Kriging can be seen as a point interpolation, which requires a point map as input and which returns a raster map with estimations and optionally an error map. Universal Kriging is a variant of the Ordinary Kriging operation: Universal Kriging is Kriging with a local trend. The local trend or drift is a continuous and slowly varying trend surface on top of which the variation to be interpolated is superimposed. The local trend is recomputed for each output pixel and the operation is therefore more similar to the Moving Surface operation than to the Trend Surface operation.

Geostatistical methods for interpolation like Universal Kriging start with the recognition that the spatial variation of any continuous attribute is often too irregular to be modeled by a simple, smooth mathematical function. Instead, the variation can be better described by a stochastic surface. The attribute is then known as a regionalized variable. The regionalized variable theory assumes second order stationarity in the data. Second order stationarity means that the mean has to exist, is constant and is independent of the location within the region of stationarity. Furthermore, the covariance or variogram has to exist, only depends on the distance between any two values, and does not depend on locations.

Universal Kriging method has formed in the Deterministic Interpolation. The assumptions spatial variation in the z together and are related to the spatial know. It is also a way to adjust the cattle of the area by integration into the sensitive surface. Quadratic planar surface (Quadratic) is the used form of a polynomial equation.

$$M = b_1 x_i + b_2 y_i$$

$$M = b_1 x_i + b_2 y_i + b_3 x_i^2 + b_4 x_i y_i + b_5 y_i^2$$

E. Analysis

1. Inverse Distance Weighting, OrdinaryKriging and Universal Kriging

In this section we give a brief description of IDW, OR and UK as well as an ad-hoc 2-step approximation to OK that can be easily fit using widely available software. We separately fit IDW, OR and UK models to the log-transformed PM_{10} concentration data from each of the three seasons. Log-transformed PM_{10} concentrations approximated a Normal distribution. Additional details of the IDW, OK, UK and the 2-step approach can be found in the Appendix. IDW, OK and UK can also be seen as a regression with geographic covariates but with the addition of correlated residuals. In other words, IDW, OK allows information about nearby concentration measurements to influence the predictions through the estimated correlation structure. Kriging is a minimum mean-squared error approach to spatial prediction [5-6].Unlike ordinary and simple kriging that assume an underlying constant mean surface, IDW, OK and UK can be easily integrated with model. The mean and covariance parameters for IDW, OK and UK are estimated together using maximum likelihood through the geoR package in R [7]. We assumed an exponential correlation structure.

ArcGIS is a widely available tool, but it does not allow IDW, OK and UK with an arbitrary set of covariates or using maximum likelihood fitting. The Inverse Distance Weighting,

Ordinary kriging and Universal kriging option within ArcGIS models the spatial correlation as expected and allows the mean to be defined as a function of the latitude and longitude, but does not allow the use of land-use covariates to be included in the mean model. The 'Inverse Distance Weighting', 'Ordinary Krigings' and 'Universal Kriging' methods in ArcGIS model the spatial correlation and fit a constant mean. With these limitations in mind we also evaluated an ad-hoc 2-step approach to prediction, first estimating the regression model assuming independent errors and then using Inverse Distance Weighting and kriging in ArcGIS to estimate the dependence in the regression residuals and exploit this dependence to improve the predictions by smoothing. This approach, which is similar but not identical to IDW, OK and UK has some advantages: it allows for spatial structure in the residuals, whereas regression does not, and it can be implemented with ArcGIS.The 2-step procedureperforms well in our examples, but we note that it may not be optimal because, unlike in IDW, OK and UK the regression coefficients are estimated without accounting for spatial correlation.

2. Cross-validation

The analysis in this paper employs a 10-fold cross-validation approach [8]. The exposure data are split into ten roughly equal size groups. Then one group of data is set aside (test set) and the model is fit on the remaining nine groups (training set). The model estimated using the training set is used to predict values at the monitor locations in the test set. This is repeated until predictions for all groups have been generated. To evaluate the predictive ability of the model, R2 and RMSE [9] are calculated based on comparing these predicted values to the observations. We refer to these as the RMSE.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i-1}^{n}(z(u)-z^*(u))^2}$$

z(x)* are interpolated unknown points at station *(i)*, *z(u)* are observed points at station *(i)*, *N* are numbers of all stations.

3. Model selection

Variable selection was done separately for each season and modeling approach (IDW, OK and UK) all starting with the same set of covariates. We excluded covariates without at least 5% nonzero values in our data. Then the least absolute shrinkage and selection operator (lasso) was implemented as a prescreening tool via the glmnet function found in the R package glmnet [10]. The tuning parameter l was selected to yield nonzero regression coefficients from the lasso. We followed the lasso algorithm with a modified exhaustive search of the possible covariate combinations. The exhaustive search approach ensures that the model selection process accounts for the spatial correlation in the data, while the lasso algorithm does not account for this correlation. The final models were selected based on predictive ability as measured by RMSE.

III. Results and Discussion

This research was conducted experiments by using ArcGis in the interpolation for accurate estimation. We can experiment by optimization parameters of each monthly data set by using IDW, OK and UK with distinct parameters.

The data of PM_{10} was selected from 30 air monitoring station in central region of Thailand. The average data of $PM_{10}$was used since January 2013 to December 2014,there are 25 stations used as training data and 5 station used as testingshown in Table 1

TABLE I
THE AVERAGE DATA OF PM_{10} FROM PCD

Parameters	Training	Testing
Station	25	5
time	Jan 2013 – Dec 2014	Jan 2013 – Dec 2014

The kriging methods between Inverse Distance Weighting (IDW), Ordinary Kriging (OK) and Universal Kriging (UK) were compared. The interpolation of PM_{10} by IDW method every month since January 2013 - December 2014 in the central regions of the country were shown in Fig.1a. The interpolations of PM_{10} by OK method and UK method at the same time period in the central regions of the country were shown in Fig.1b and 1c, respectively.

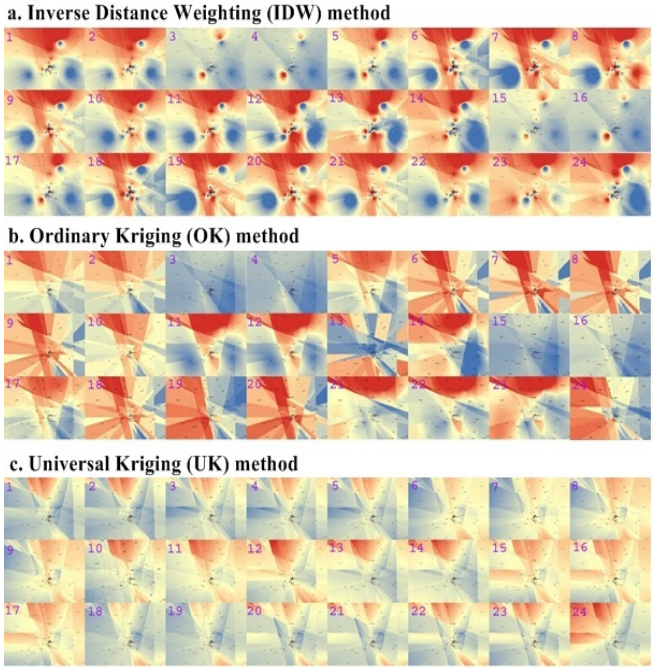


Fig. 1The interpolation of PM_{10} by IDW method (a), OK method (b) and UK method (c) by mouth since January 2013 - December 2014 in the central regions of the country

From Fig2, there are the results from interpolation testing data of PM_{10} with all method from 5 airs monitoring stations and comparing the prediction performance (for further information see the appendix). The 5 air monitoring are Post Office Ratburana station (S1), HuaiKhwang station (S2), Chulalongkorn University station (S3), Public Relations Department station (S4) and PrabadangRehabitation for the Disable station (S5)

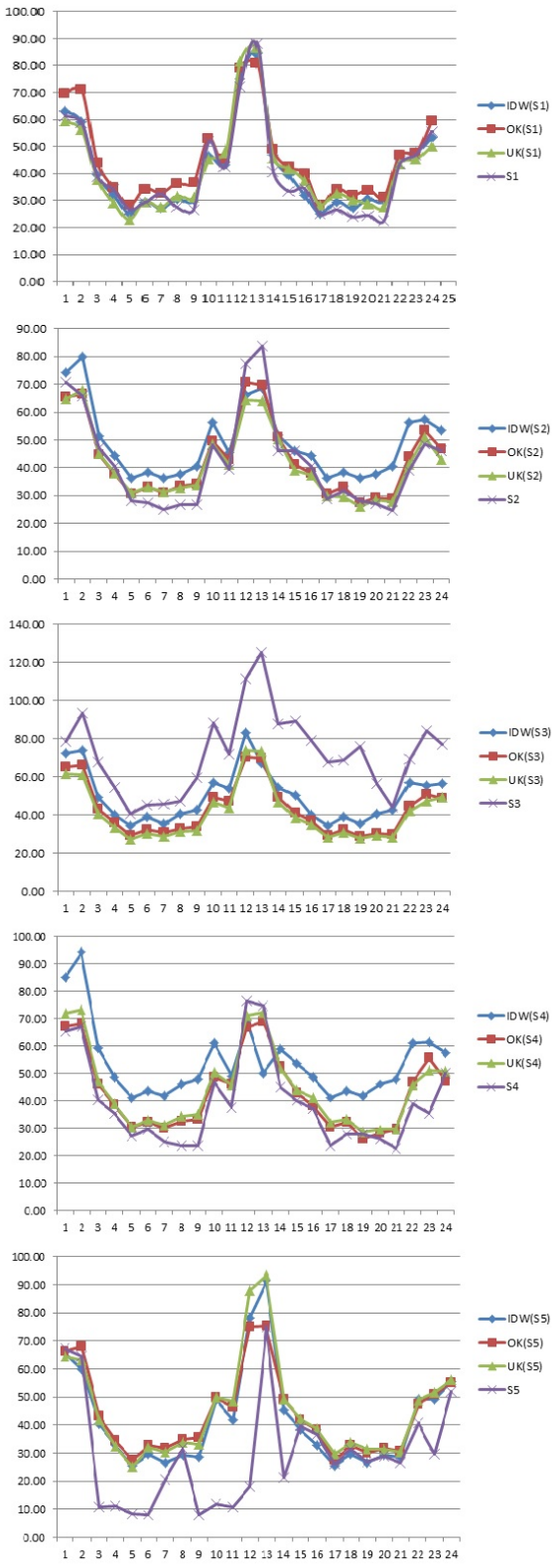


Fig. 2 The results from interpolation testing data of PM_{10} with all method from 5 air monitoring stations and comparing the prediction performance

Root Means Square Error value from the every station predicted by interpolate with IDW method, OK method and UK method shown in Table 2 follow by graph shown the relation of RMSE of each interpolated method (Fig.3)

TABLE II
ROOT MEANS SQUARE ERROR VALUE FROM EACH INTERPOLATED METHOD

	RMSE		
Station	IDW	OK	UK
1	0.703	0.715	0.714
2	0.684	0.695	0.71
3	0.683	0.693	0.714
4	0.687	0.711	0.721
5	0.694	0.711	0.722

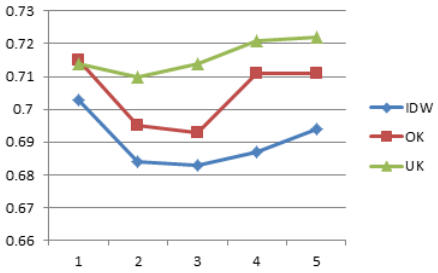


Fig. 3 The relation of RMSE of each interpolated method

IV. CONCLUSION

The prediction of PM_{10} by interpolating the Inverse Distance Weighting (IDW) method, Ordinary kriging (OK) method and Universal Kriging (UK) method. The comparison between each method found that Inverse Distance Weighting (IDW) method is the best interpolated method to predict the air pollution.

APPENDIX

Supplementary material for the results from interpolation testing data of PM_{10} with all method from 5 airs monitoring stations can view in

https://drive.google.com/file/d/0BxDtk22BGcMidGRRWldpU0FOQU0/view?usp=sharing

ACKNOWLEDGMENT

This study was supported by BangkokThonburi University, Thailand.

REFERENCES

[1] L. Xian, et al., "Comparison of interpolation methods for spatial precipitation under diverse orographic effects," in Geoinformatics, 2011 19th International Conference on, pp. 1-5, 2011.
[2] A. Mair and A. Fares, "Comparison of Rainfall Interpolation Methods in a Mountainous Region of a Tropical Island," Journal of Hydrologic Engineering, vol. 16, pp. 371-383, 2011.
[3] J. Li and A. D. Heap, "A Review of Spatial Interpolation Methods forEnvironmental Scientists", Geoscience Australia, Canberra, 2008.
[4] F. W. Chen and C. W. Lui, "Estimation of the spatial rainfall distribution using inverse distance weighting (IDW) in the middle of Taiwan," the Paddy and Water Environment, vol.10, pp.209-222, February 2012.
[5] S. Banerjee, B. P. Carlin and A. E. Gelfand, Hierarchical Modeling and Analysis for Spatial Data. Chapman and Hall, UK, 2004.
[6] N. A. C. Cressie, Statistics for Spatial Data. Wiley Series in Probability and Mathematical Statistics. Wiley-Interscience, NJ, 1993.

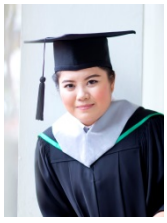
[7] P. J. Ribeiro and P. J. Diggle, "geoR: A package for geostatistical analysis." R-NEWS, vol.1, June 2001.
[8] T. Hastie, R. Tibshirani and J. Friedman, The Elements of Statistical Learning Data Mining, Inference, and Prediction. Springer, Germany, 2001.
[9] Y. Sun, et al., "Comparison of interpolation methods for depth to groundwater and its temporal and spatial variations in the Minqin oasis of northwest China," Environmental Modelling & Software, vol. 24, pp. 1163-1170, 2009.
[10] J. Friedman, T. Hastie and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," Journal of Statistical Software, vol.33, pp. 1-22, Aug 2010.

Phatarapon Vorapracha (M.Sc.) was born in Nakhon Pathom province, Thailand on July 25, 1982. He finished Master's degree in Information Technology from Faculty of Science, Silpakorn University, Nakhon Pathom, Thailand in 2008 and Bachelor's degree in Computer Science from Faculty of Science, Nakhon Pathom Rajabhat University, Nakhon Pathom, Thailand in 2003.

He used to work as Programmer at Urmet (Thailand) Co.,Ltd. After that, he was teacher at Nakprasith School and a systems analyst at Alloy Industry Co.,Ltd. And now, he work as lecturer at Division of Information Technology, Faculty of Science and Technology, BangkokThonburi University, Bangkok, Thailand.

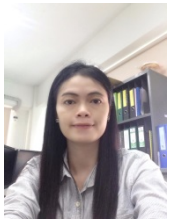
Pongtep Phonprasert (M.Sc.) was born in Nakhon Pathom province on January23, 1984. He finished Master's degree in Environmental Science and Technology from Faculty of Science, Kasetsart University, Nakhon Pathom, Thailand in 2012 and Bachelor's degree in Environmental Science from Faculty of Science, Suan Dusit Rajabhat University, Bangkok, Thailand in 2005.

He used to work as Environment Scientist at Envi-Pro Co, Ltd. After that, he was Leader of Environmental Monitoring Team, Green Tea Environment Services Co, Ltd. (Siam Cement Group's subsidiary company), Bangkok, Thailand. In the present, he works as lecturer at Division of Environmental Technology, Faculty of Science and Technology, BangkokThonburi University, Bangkok, Thailand.



Suparada Khanaruksombat (M.Sc.)was born in Ratchaburi province on February20, 1988. She finished her Master's degree in Bioproduct Science from Faculty of Liberal Arts and Science, Kasetsart University, Nakhon Pathom, Thailand in 2014 and graduated the Bachelor's degree in Biological Science from Faculty of Liberal Arts and Science, Kasetsart University, Nakhon Pathom, Thailand in 2010.

In the same time on her Master's degree study, she was a research student in Biochemistry Laboratory, Chulabhorn Research Institute, Thailand.In the present, she works as lecturer at Division of Environmental Technology, Faculty of Science and Technology, BangkokThonburi University, Bangkok, Thailand.



Nuchanaporn Pijarn (Ph.D.) was born in Ubon Ratchathani province on 15 April 1978. Doctor of Philosophy was received in Materials Science, Faculty of Science, Chulalongkorn University, Bangkok, Thailand in 2013. Master's degree was received in Applied Analytical and Inorganic Chemistry, Faculty of Science, Mahidol University, Bangkok, Thailand in 2007. Bachelor's degree was received in Chemistry, Faculty of Science, Ubon Ratchathani University, Ubon Ratchathani, Thailand in 2001.

She is currently a lecturer at Division of Environmental Technology, Faculty of Science and Technology, BangkokThonburi University, Bangkok, Thailand. She was assistant-researcher at Nation Metal and Materials Technology (MTEC), National Science and Technology Development Agency (NSTDA), Pathumthani, Thailand, and a lecturer at department of Chemistry, Faculty of Science, Ubon Ratchathani University.