2016 International Conference on Materials, Information, Mechanical, Electronic and Computer Engineering (MIMECE 2016)
ISBN: 978-1-60595-420-2

Research on Online Mining Method of Real Time Soil Sensing Data in Farmland

Cheng Peng, Huarui Wu, Ronghua Gao and Yuansheng Wang

ABSTRACT

With the wide application of various agriculture IOT sensor equipment, sensor networks produce ample data for observing dynamic soil properties in farmland. But data processing for analysis and visualization become increasingly difficult as data dimensionality increases. Meanwhile the devices sometimes lose data because of damage and power outages. In this study, exploratory data analysis (EDA) and spatio-temporal interpolation including time as a third dimension and spatial-temporal variogram are used to analyze the soil moisture and temperature data. The experimental results show that EDA can effectively reveal the implicit information of soil sensing data in multiple perspectives by a variety of visual ways. The leave-one-station-out cross-validation indicate that Kriging method based on spatial-temporal variogram has satisfied accuracy, which supplies an effective approach for interpolation and estimation of lost agriculture IOT sensing data. The results also have theoretic and practical value for regional soil management.

INTRODUCTION

With the rapid development of Internet of Things, wireless communication technology and the popularize of the "Internet plus agriculture", more and more large farm, agricultural zones and farmer specialty cooperative organizations apply IOT technology to improve the level of intelligence, precision and automation of agricultural production. The radio frequency identification device, global positioning system and other various sensing devices are used to capture agricultural production site information, including soil temperature and water content, air temperature and humidity, light illumination intensity, soil element content, geographic location information and environmental information data, which ensure the digitization management of agricultural production.

Cheng Peng, Huarui Wu*(corresponding author), Ronghua Gao, Yuansheng Wang. National Engineering Research Center for Information Technolgoy in Agriculture, Beijing Nongke Mansion, No 11 Shuguang Huayuan Middle Road, Banjing, Haidian district, Beijing, P.R. China E-mail: wuhuarui1975@163.com

These sensing devices provide high frequently updated measurements of environmental properties at fixed locations, including 2- or 3- dimensional and temporal data. Although these devices collected enough data to measure the dynamic soil properties, the visualization and analysis on multidimensional data temporal data. Although these devices collected enough data to measure the dynamic soil properties, the visualization and analysis on multidimensional data become more and more difficult with the continuous accumulation of data amount.

Exploratory data analysis (EDA) is an approach for data analysis to summarize their main characteristics and explore their structures and laws. Different from the traditional statistical analysis method, EDA does not need to have a hypothesis, priori knowledge and verification, but gradually generate hypotheses and seek its answer in the process of data analysis. The implicit information of the massive and continuous soil sensing data can be aggregated, abstracted, discovered and extracted by using EDA.

On the other hand, in the long time period of sensor's monitoring, the agricultural IOT sensing equipment often suffer the power outages, broken network or hardware damaged, resulting in loss of data. Multiple approaches have been developed for spatial interpolation of sensing data, including: (1) multiple regression models based on the forming factors, terrain attributes, spatial coordinates, or derived principal components [1]; (2) geostatistics, or Kriging, and variations [2-5].

The soil sensing data in farm belong to spatial and temporal data, which usually refers to the continuous and discrete data of time and space. The observation data of space-time elements not only has a strong spatial correlation in space, but also has the characteristic of time series in time. Continuous surface can be obtained by spatial interpolation method and missing data are filled by time series method. Nevertheless, the traditional spatial interpolation method often destroy the unity of space-time continuum and make the time and space analysis confined to a specific time point or time. Moreover, the simple time series analysis ignores the spatial correlation in geographical space. Therefore these two methods do not take into account complex temporal and spatial relationship of spatio-temporal data and cause the loss of dimension information of spatio-temporal data. As a result, analysis on spatial and temporal data in view of single time or space leads to the loss of valuable information, which is not suitable for the improvement of the interpolation precision.

Spatio-temproal Kriging method extend Kriging from the space and time domain, which makes full use of time information of data in the process of interpolation. Recent studies have shown that spatial-temporal Kriging method has applied in a range of diverse fields such as regional geomagnetic field estimation, meteorological element interpolation, and distribution of soil heavy metal, soil salinity and wind data interpolation and so on. Gasch et al. [6] took hourly measurements of soil volumetric water content, temperature, and bulk electrical conductivity at 42 stations and five depths as example and proposed two approaches to produce continuous predictions from 3D + T point observations of three dynamic soil variables by a 3D sensor network for multiple years. Aryaputera et al. [7] performed very short-term irradiance forecasting using a dense solar irradiance monitoring network applying several variants of spatiotemporal Kriging. Lisha et al. [8] constructed a kind of product-sum variogram in space-time to describe the spatial-temporal correlation based on pure spatial variogram and pure temporal one. To our knowledge spatio-temporal Kriging methods have not yet been expanded to produce predictions from data collected in soil sensing data of agricultural IOT.

The aim of this study was to introduce exploratory data analysis and spatio-temporal interpolation method for mining continuous and ample soil content variables obtained at different IOT sensor stations over the farm. The rest of the paper is organized as follows. Section 2 explains the data from the farm sensing network and analytical methods. Section 3 displays the exploratory analysis and interpolation results. Finally, Section 4 presents the conclusion.

MATERIALS AND METHODS

Data Set

The soil sensing data used in this paper are taken from the agricultural IOT monitoring network in Huitian farm. The study area covers about 1.8 km^2 and is located in Fangshan District of Beijing (39° 43′ N, 116° 6′ E) in the northeast of China. The locations of the sampling sites are shown in Figure 1. From each station, the soil moisture (%) and temperature (°C) in four depths (0.2, 0.4, 0.6, 0.8m) at an hourly rate are available. The data cover a time period from January 1 in 2015 to December 31 in 2015.

Figure 1. Location of 18 soil sensing stations in Huitian farm in Fangshan District of Beijing.

Exploratory Data Analysis

EDA employs a variety of techniques involves in extracting important variables, test underlying assumptions and determining optimal factor setting. EDA often uses a variety of visualization techniques to uncover underlying deep information and make data more easily understood. There are a number of tools that are useful for EDA such as histogram, scatter plot, multidimensional scaling, and targeted projection pursuit and so on.

Spatio-temporal Kriging Method

Kriging method was first developed by Matheron [9] based on the work of Krige, a mining engineer from South Africa. Kriging interpolation is a linear unbiased estimation method, which requires the minimum variance of estimation error. It interpolated the estimation point according to the spatial autocorrelation of variables based on the covariance (variation) function of the variables. Kriging has later been generalized for spatio-temporal applications [10].

(1) Time as the third dimension

The equation to calculate the sample semivariogram is as bellows [11]:

$$\hat{r}(h) = \frac{1}{2N(h)} \sum_{N(h)} \{[Z(s_i) - Z(s_j)]\}^2 \quad (1)$$

In Equation (1), the vector h is the distance between two points s_i and s_j, which determines their variance. It is important to note that the temporal dimension needs to have a range similar to the spatial dimension; as a result time needs to be scaled to accord with the spatial dimension.

(2) Spatio-Temporal Variogram

The second way of taking time into account is to adapt the covariance function to the time component. The variance between point s_i and another is calculated associated with a time t_i. In Equation (2), h is their spatial separation and u is their temporal separation. Thus, the spatio-temporal variogram can be computed as follows [11]:

$$\hat{r}(h,u) = \frac{1}{2N(h,u)} \sum_{N(h,u)} \{[Z(s_i,t_i) - Z(s_j,t_j)]\}^2 \quad (2)$$

Software Implementation

All analysis was performed in R. The packages of aqp, gdata, rattle, ellipse, rgdal, raster and etc. were utilized to process soil sensing data. The Kriging approach was mainly based on the gstat package [12] in combination with the spacetime package [13]. The package spacetime has classes and methods for spatio-temporal and provides ways of creating objects where the time component is taken into account. Then these formats are used by gstat for its space-time analysis. Gstat is a program for the modelling, prediction and simulation of geostatistical data in one, two or three dimensions. Rattle provides a graphical user interface to very many other R packages that involve functionality for data mining .

RESULTS AND DISCUSSION

Descriptive Statistics

The annual average value of soil moisture for 4 depths in 19 monitoring points was calculated by traditional statistical routine firstly. As it shown in Table 1, the average soil moisture values in each depth had moderate variation. Moreover, the other statistical values were related to the mean and reflect the overall distribution only in a certain extent. Thus it is necessary to characterize the randomness, irregularity, independence and correlation of soil moisture using exploratory data analysis and spatial analysis.

Table 1. Statistical characteristics of soil water content at different depths.

Soil depth (CM)	Mean value	SE Mean	LCL Mean	Variance	Standard deviation
20	21.46	0.0029	21.41	47.57	6.897
40	22.23	0.0028	22.17	45.48	6.744
60	25.15	0.0029	25.09	49.63	7.045
80	22.8	0.0311	22.74	54.63	7.391

Exploratory Data Analysis

Investigating the correlation between each variable is one of the processes to make a preliminary understanding of data set. There are 10 variables including soil temperature and humidity in four different depths, as well as air temperature and rainfall. STC20, STC40, STC60 and STC80 respectively represent soil temperature in depth of 20cm, 40cm, 60cm and 80cm; similarly, SWC20, SWC40, SWC60 and SWC80 respectively represent soil moisture in 4 different depths. ATC is representative of air temperature and AWS is on behalf of precipitation.

The relevant figure of the 10 variables is such as Figure 2. The width of circular reflect the correlation of two variables, like that the circular narrower, the correlation higher. In addition, direction of the inclined circular is related to the positive or negative correlation. The right incline means positive correlation. It can be seen from Figure 2 that the soil moisture of each depth has a little correlation as well as the temperature and humidity between soil and air. But apparently there is a strong correlation among soil temperatures in different depths.

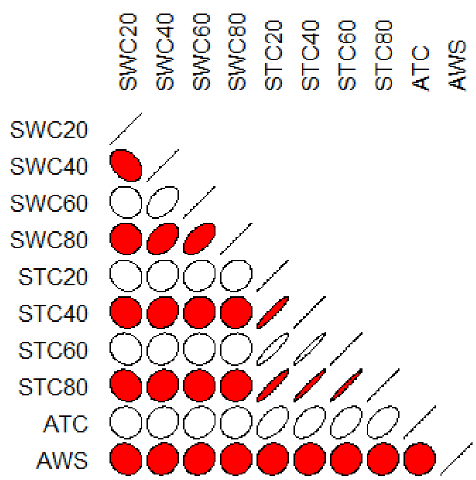
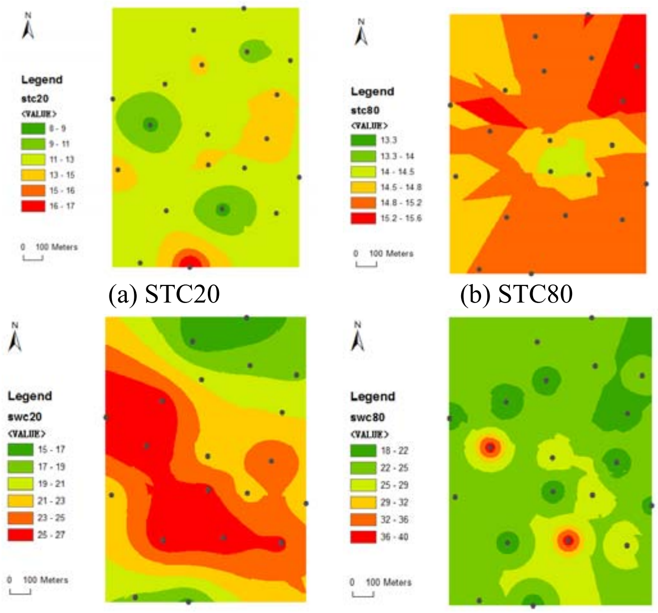


Figure 2. The relevant figure of 10 variables on temperature and humidity of the soil and air.

Spatial Kriging

The Kriging map of average soil moisture and temperature in October 2015 can be seen from Figure 3, which showed the distribution of soil properties in farm accurately and visually.



(a) STC20 (b) STC80

(c) SWC20 (d) SWC80

Figure 3. Map of soil moisture and temperature in 20CM and 80CM depth.

Comparison of Spatio-temporal Kriging and Other Interpolation Methods

In order to test the actuary of 4 methods including time series, nearest neighborhood, ordinary Kriging and spatio-temporal Kriging, the leave-one-station-out cross-validation were run separately. It took at a certain station as unknown in each of the 19 stations [14] and $\hat{Z}_i$ was calculated as the values of the soil temperature and humidity at station i by values of the other 18 stations, then record the error between the measured values and the calculated value $\left|Z_i - \hat{Z}_i\right|$. Each station is used as a position point, i=1:19. The evaluation standard is the Mean Absolute Error and Mean Square Error between the predicted and measured values.

$$MAE = \frac{1}{n}\sum_{i=1}^{19}\left|Z^*(s_i,t) - Z(s,t)\right| \tag{3}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{19}(Z^*(s_i,t) - Z(s,t))^2 \tag{4}$$

The time sequence length is 365 days. 30 days before one day t [t-30,t-1] were taken as space-time database and interpolation of cross validation were implemented by above 4 methods. The results are shown in Table 2. No matter MAE or MSE, the error of time series, nearest neighbor and ordinary Kriging are larger than that of spatio-temporal Kriging, whose results are more stable and the error are small. MAE is concentrated in about 1 and the concentration of MSE is in the range of 0, 5.

Table 2. Error statistics results of 4 interpolation methods.

Method	Mean of MAES	Mean of MSES	Variance of MAES	Variance of MSES
Time series	1.9456	6.9521	0.1836	15.3698
Nearest	1.7599	6.2583	0.1987	18.1452
Ordinary	1.5351	3.8677	0.0929	3.1577
Spatio-temporal	1.0328	2.1993	0.084	3.3982

CONCLUSIONS

With the widespread application of agricultural IOT, distributed sensor networks provide frequent measurements of environmental properties at fixed locations, providing data in 2- or 3-dimensions and through time. But the sensing data may be lost with the damage of hardware or software. Ideally, the end product should consist of seamless interpolations that accurately represent the spatial and temporal variability in the property of interest. These products can then be used for predictions at unobserved locations and can be integrated into process models, and they can simply aid in visualization of soil properties through space and time.

In this paper, the original data were analyzed by EDA firstly for exploring the soil moisture and temperature distribution characteristics overall farm. On the basis of descriptive statistics, the spatial characteristics in the representative time period were discussed. The process generates new problems and continues to analyze the progressive reasoning, fully reflecting the characteristics of data and the nature of the data. It probes into the implicit information of soil sensing data from general to individual, from detailed to classified, from static to dynamic.

Then time and space domains were expanded with the effect of method and the time domain information was added. Compared with the conventional time series, nearest neighborhood and ordinary interpolation methods, the cross validation error of spatio-temporal Kriging method decreases and the interpolation accuracy improves. These methods are available and practical for the exploration and prediction of continuous observation on soil moisture and temperature in agricultural Internet of things.

ACKNOWLEDGEMENTS

This work was supported by Beijing Municipal Natural Science Foundation (4151001) and National Natural Science Foundation of China (41501418).

REFERENCES

1. McKenzie, N.J., Ryan, P.J. 1999. “Spatial prediction of soil properties using environmental correlation”, *Geoderma*, 89(1-2): 67-94.
2. McBratney, A.B., Mendonça Santos, M.L., Minasny, B. 2003. “On digital soil mapping”, *Geoderma*, 117(1-2): 3-52.
3. Hengl, T. 2009. *A Practical Guide to Geostatistical Mapping.* Lulu.com, Amsterdam, Netherlands, pp: 44-55.
4. Hegedus. P., Czigany. S., Pirkhoffer. E., Balatonyi. L. and Hickey, R. 2015. “Analysis of spatial variability of near-surface soil moisture to increase rainfall-runoff modelling accuracy in SW Hungary”, *Open Geosciences*, 7(1): 126–139.
5. Cichota, R., van Lier, Q.D. 2004. “Spatial variability analysis of sampling points of the soil water retention curve”, *Revista Brasileira De Ciencia Do Solo*, 28(4): 585–596.
6. Caley K. Gascha, Tomislav Hengl B., Benedikt Gräler, Hanna Meyerd, Troy S. Magneye, David J. Browna. 2015. “Spatio-temporal interpolation of soil water, temperature, and electrical conductivity in 3D + T: The Cook Agronomy Farm data set”, *Spatial Statistics*, 14(4): 70-90.

7. Aloysius W. Aryaputera A,B, Dazhi Yang, Lu Zhao, Wilfred M. Walsh. 2015. “Very short-term irradiance forecasting at unobserved locations using spatio-temporal”, *Solar Energy*, 122: 1266–1278.
8. Li S., Shu H., Xu Z.Q. 2012. “Interpolation of Temperature Based on Spatial-temporal Kriging”, *Geomatics and Information Scienc of Wuhan University*, 37(2): 237-241.
9. Matheron, G., 1963. *Principles of Geostatistics*. Econ. Geol. 58, 1246-1266.
10. Noel Cressie and Christopher K. Wikle. 2011. *Statistics for spatio-temporal data*. Wiley, pp. 89-112.
11. Sherman, M., 2011. Spatial statistics and spatio-temporal data: covariance functions and directional properties. John Wiley & Sons, pp. 123-132.
12. Pebesma, E., Gräler, B., 2013. “gstat: spatial and spatio-temporal geostatistical modelling, prediction and simulation”. R package version 1.0-16. URL: http://CRAN.R-project.org/package=gstat.
13. Pebesma, E., Bivand, R., 2013. “sp: classes and methods for spatial data. R package version 1.0-5. URL: http://CRAN.Rproject.org/package=sp.
14. Laurenceau J., Sagaut P. 2008. “Building efficient response surfaces of aerodynamic functions with Kriging and Cokriging ”. *AIAA Journal*, 46(2): 498-507.