Methods to define confidence intervals for kriged values: Application on Precision Viticulture data.

J-N. Paoli[1], B. Tisseyre[1], O. Strauss[2], J-M. Roger[3]
[1]*Agricultural Engineering University of Montpellier, FRANCE*
[2]*LIRMM Montpellier, FRANCE*
[3]*CEMAGREF Montpellier, FRANCE*
Email: paoli@ensam.inra.fr

Abstract

In Precision Viticulture, the increasing number of information sources calls for methods to combine them and to generate information for professionals. Each data is located on its own geographical support, and geostatistics are mostly used to produce estimates on a common grid. Estimated values have to be associated with confidence intervals; using kriging variance to compute them is generally considered to be efficient. This paper discusses this approach and presents two original methods to compute confidence intervals without using kriging variance. These methods are based on iterations of the kriging step, and takes into account inaccuracy of variogram estimation or measurement variance. They have been applied to yield data. Results have been compared to those obtained with kriging variance.

Keywords: Precision Viticulture, geostatistics, confidence intervals, variogram cloud, bootstrap

Introduction

Professionals (such as farmers or consultants) who work in the field of agriculture and more specifically of viticulture will have access to an increasing number of information sources (yield maps, sugar rate maps, remote sensing images…). The precision and the resolution of these data are variable, and the major problem is to combine them in order to define homogeneous zones within fields or to compute data that could be used by professionals, such as "potentially qualitative zones of vintage" (Bramley, 2001; Tisseyre et al, 2001).

Because data are usually neither similarly nor regularly distributed on the map, computation of homogeneity criteria becomes intricate. To overcome this problem, geostatistics and kriging methods are extensively used to transform irregularly to regularly distributed data. A common grid is defined and data are interpolated to provide an estimated value on each node of the grid for each considered source. It is, now, essential to characterize the quality of this new kriged map. Users usually admit that kriging variance could be used to compute an estimate of the standard deviation of the kriged data. This approach has to be discussed.

Kriging variance depends on the variogram and on the configuration of sampling design (Journel and Huigbregts, 1978; Arnaud and Emery, 2000). With high-resolution data (such as yield data), kriging variance is close to nugget effect, which can be observed in the variogram. However, kriging variance doesn't take into account inaccuracy of variogram estimation, which could widely influence the estimation process. Moreover, the choice of the function that is used to model the variogram can bias the estimated value of the variance.

It is clear that the kriging procedure is based on the underlying hypothesis that measurement error is uniformly distributed on the field. Although, a variogram computed on the half of a field can be significantly different from the one computed on the other half. Usually, this difference is given to be due to computation or sampling noise. With intensive data sets, block kriging could be applied in order to eliminate this noise. But this difference can also be explained by local variations of measurement error. To discard this information can alter the sough after segmentation of the field into different homogeneous regions. We aim to compute this information to give the accuracy of the kriging procedure, according to local measurement variances.

This work presents two methods that take into account inaccuracy of variogram estimation and variations of measurement error. These two methods are based on iterations of an ordinary kriging procedure, and provide a variance for each estimated value on the grid. The values and variances provided by our methods are compared to those one give by "ordinary kriging" method.

Materials and Method

Experimental field

Our experimental vineyard is an area of Syrah variety trained in Royat cordon. Stocks are 8 years old (density of 4000 stocks ha^{-1}), and trained to a height of 1.7 m. This area of 1.2 ha is on the "Clape limestone massif". Grape yield are measured in 2001 using an on-line sensor mounted on a grape-harvesting machine (Pellenc S.A.). The generated database includes 2400 yield values. Figure 1 shows the yield variability measured by the grape-harvesting machine.

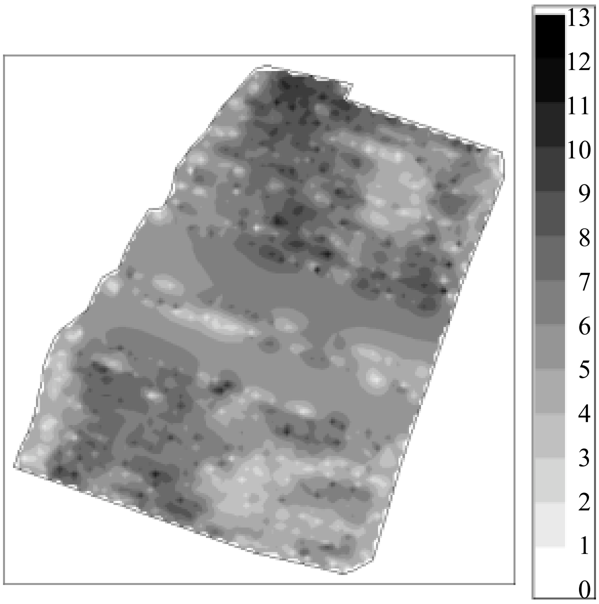

Figure 1. Grape yield map ($t.ha^{-1}$), first view of the yield with a simple interpolation method (Inverse Distance weighting)

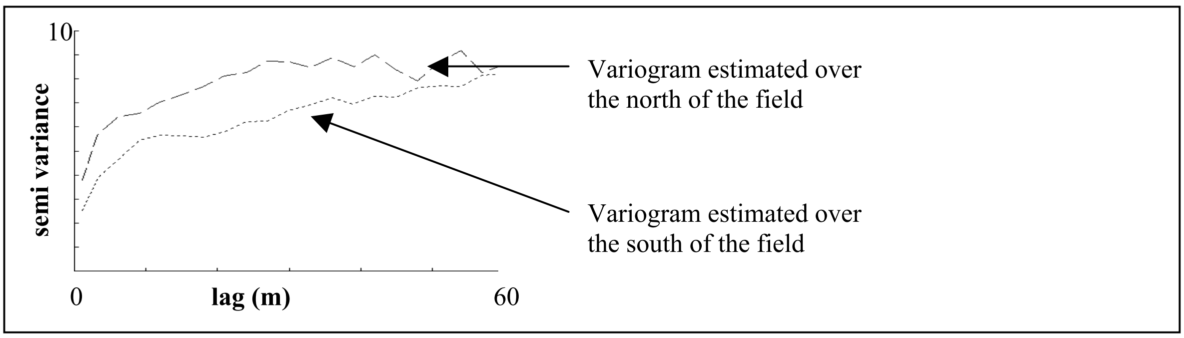

Figure 2. Local experimental variograms

In a first experiment, we have divided the field in two parts (north and south). Figure 2. shows that there is a significant difference between the variogram computed with the north field and the variogram computed with the south field.

Test protocol
The database was divided into two sets. 600 points were randomly assigned to the training set while the remaining 1800 points were used to build a test set. The random assignment process consists in dividing each row into groups of four successive measurements and to randomly select the training point. (Figure 3).

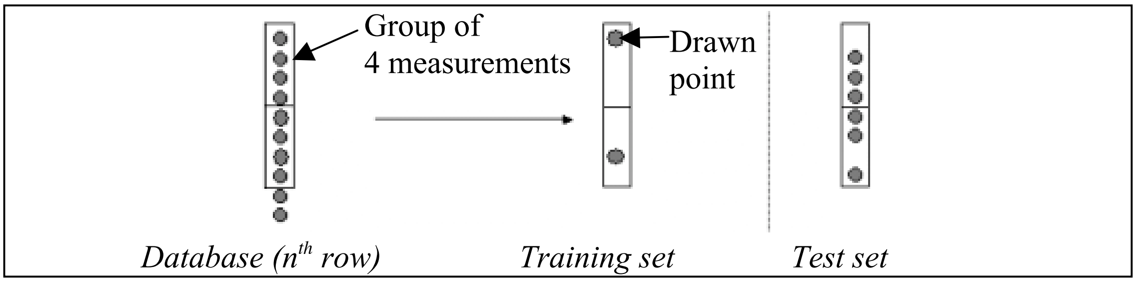

Figure 3. Division of the database into a test set and a training set

Training points are used to estimate yield data and standard deviation on each point of the test set. The estimation is performed using three different processes : "ordinary kriging"-, "variogram cloud"-, and "bootstrap" method.

The comparison of the accuracy of those three methods is done in two steps :

- The first test consists in comparing the estimation of the yield with the real data of the test base. This comparison uses the computation of Root Mean Squared Errors (RMSE) for each method.
- The second test, aimed at verifying that the estimated variance was correlated to the error between estimated values and real values. To perform this test, the test set is divided into several groups with homogeneous predicted variance. Then, we compute RMSE for each group and compare it to the predicted variance.

Variogram cloud method
This method aims to take into account the inaccuracy of the variogram estimation.

The variogram cloud (Figure 4a) is a scatter plot of the individual semi variance value for each pair of points against lag distance, and shows the spread of values at different lags. For each lag, we compute upper and lower quartiles of this cloud to obtain an upper and lower variogram. We use a parametric function (1) to model each variogram. $(C_{0-0.25}, C_{1-0.25})$ are the parameters of the function for the lower quartile and $(C_{0-0.75}, C_{1-0.75})$ are the parameters for the upper quartile (figure 4b).

$$f(d) = C_1(1 - e^{-\frac{d}{r}}) + C_0 \text{ (exponential model)} \qquad (1)$$

Any variogram modeled by the function (1) can be provided by randomly selecting the values of C_0 and C_1 with

$$C_{0,0.25} \le C_0 \le C_{0,0.75},\ C_{1,0.25} \le C_1 \le C_{1,0.75} \text{ (and } C_1 \ge C_0\text{)},\ r = 26 \text{ (fixed)}$$

We have generated 100 different values of C_0 and C_1. We then used each function for estimating the kriged values of the test set.

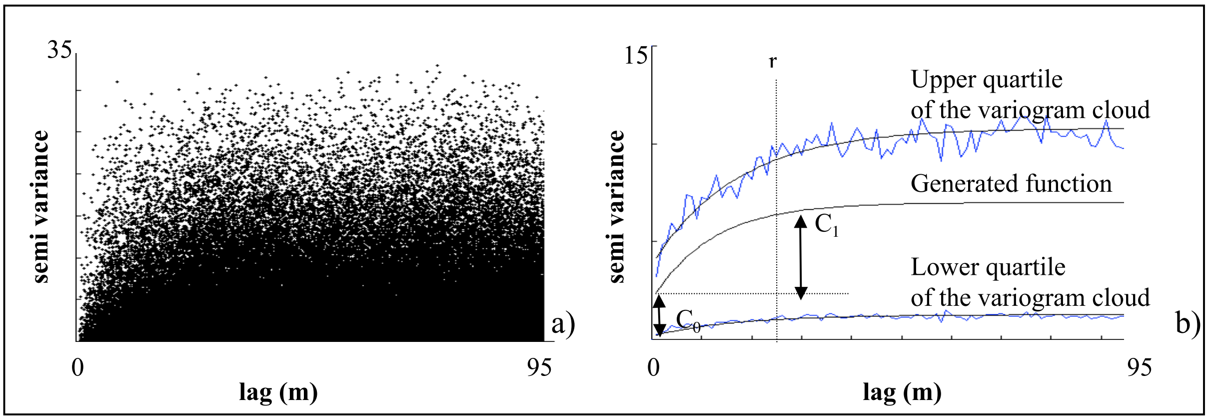

Figure 4. a) variogram cloud, b) Modeling step: Generation of several variograms, according to the variogram cloud

Bootstrap method

This method aims to give the accuracy of the kriging procedure, according to local measurement variances.

In (Whelan et al, 2001), the authors propose to compute a local variogram on the neighborhood of each estimated point. Nugget effects given by each local variogram should estimate measurement variance. However, the model given by such a method is too complex and requires specific software such as VESPER (Whelan et al, 2001).

If a single model is considered as representative enough of all the parts of the field, a method to test the estimation process could be useful. “Bootstrap” method provides such kind of test (Lecoutre and Tassi, 1987). This method is based on iterated random sub-sampling. All sub samples are roughly similar, and differences between them is attributed to measurement variance.

One hundred different sub-samples were drawn using the Bootstrap method (figure 2). Then, the kriging procedure was repeated using:

- the sub-samples
- the variogram computed from the points of the training set and used for ordinary kriging.

Results and Discussion

The RMSE computed from all the points are similar for all the tested methods (ordinary kriging method, variogram cloud method, bootstrap method). So, our methods don’t influence the overall accuracy of the kriging procedure.

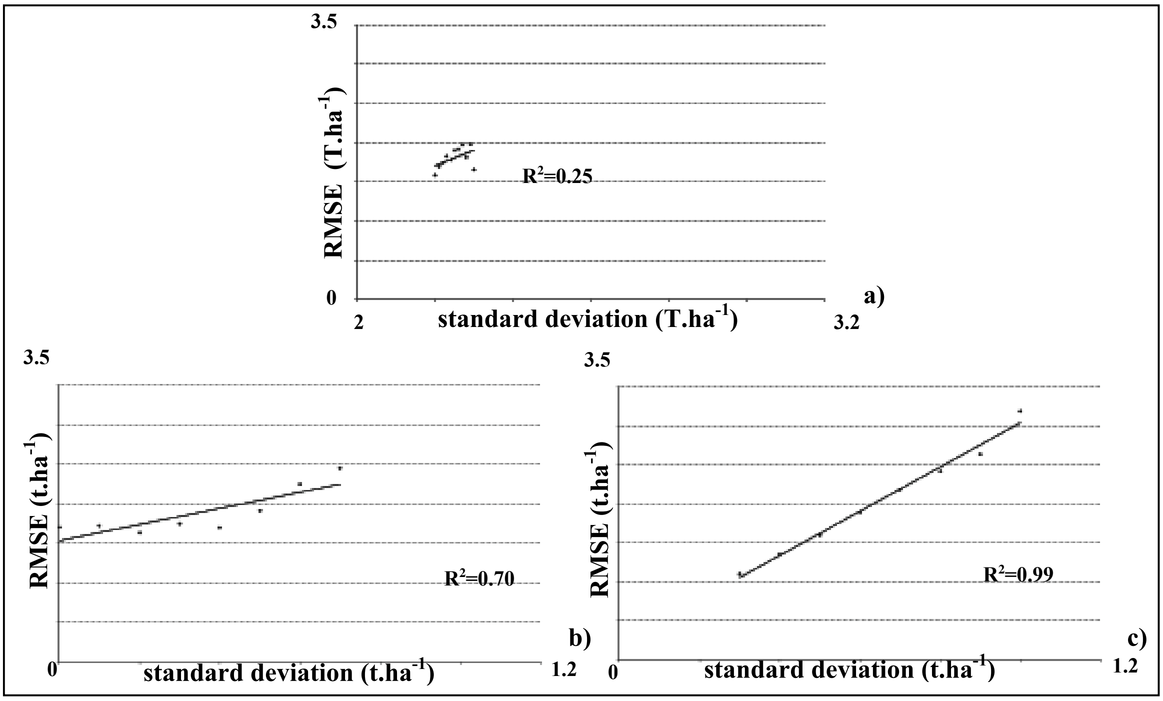

Figure 5. Correlation between RMSE and standard deviations estimated by
a) the ordinary kriging method, b) the variogram cloud method, c) the Bootstrap method

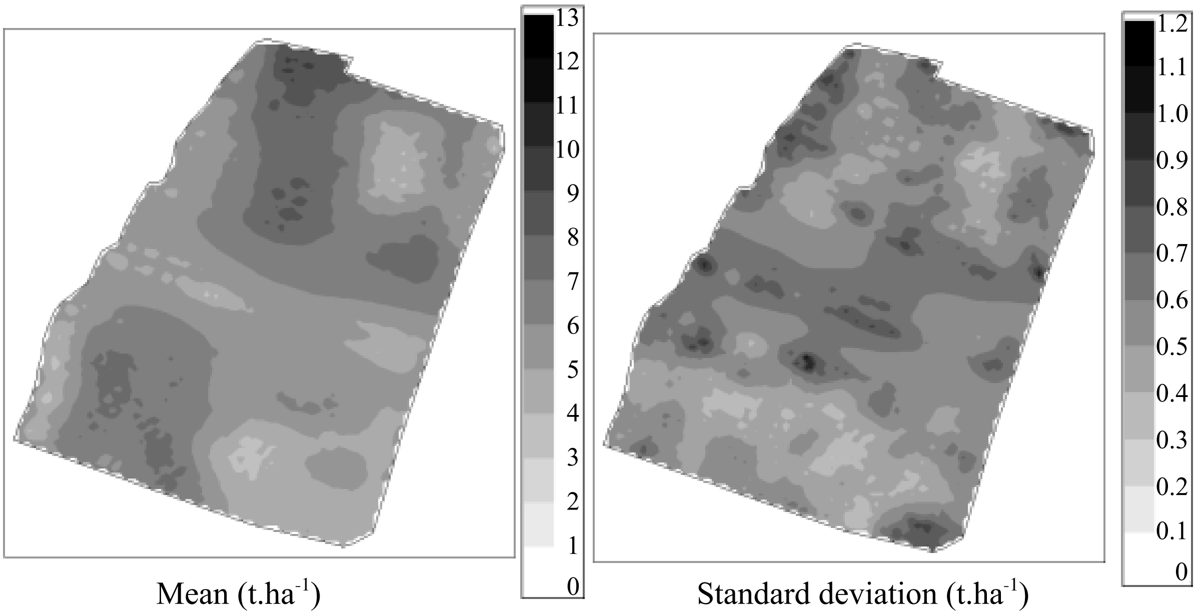

Figure 6. Method 2 -Yield mean and standard deviation Maps (interpolated)

Figure 5 shows for each method the relation between standard deviations and RMSE. On figure 5a, each graph point describes a group of estimated values characterized by homogeneous kriging standard deviations and heterogeneous squared errors. The result is a poor correlation between kriging standard deviation and RMSE. In our study case, standard deviations obtained by the variogram cloud method (Figure 5b) and by the bootstrap method (Figure 5c) are better correlated with RMSE.

The best correlation is given by the Bootstrap method. Figure 5c shows a linear relation between a standard deviation value and the RMSE computed from all the points characterized by this value. Moreover, maps of means and standard deviations (figure 6) are spatially consistent. The mean map is smoother than the yield map, but different yield zones are correctly defined.

The standard deviation map highlights transition zones and homogeneous areas. The variogram cloud method seems to be less efficient. This method should be tested on "medium resolution" data (100-200 points/ha), which may generate very inaccurate variograms. But applying the method on this type of data may require improvements of the modeling step (choice of the percentiles, of the parametric functions…).

Conclusion

In order to combine heterogeneous data (such as Precision Viticulture data), geostatistics can be used to estimate each data onto a common grid. But the accuracy of this estimation has to be evaluated and the use of kriging variance is not efficient on high-resolution data. To solve this problem, two original methods, based on iterations of the kriging step, have been developed. The first uses different variograms estimated from the variogram cloud. The second is based on a bootstrap approach. Results of iterations are used to compute means and standard deviations for each estimated point. Results were compared to those given by the classical kriging procedure on precision viticulture data sets.

Based on our data, these two methods show promising results. The first one should be developed to be applied to medium resolution data. The second one could be used to treat high-resolution data before combining them. Our future work is dedicated to the use of those estimated values to provide a segmentation of the field into homogeneous regions. Our methods will allow definition of a region growing segmentation process that could use estimation of uncertainty.

References

Arnaud, M.,Emery, X., 2000, Estimation et interpolation spatiale, méthodes déterministes et méthodes géostatistiques, (Spatial estimation and interpolation, deterministic and geostatistic methods) Hermes Science Publications eds, Paris, France, 221 p

Bramley, R.G.V., 2001, A protocol for winegrape yield maps. In: Proceedings of 3rd ECPA, edited by S. Blackmore, G. Grenier, Agro Montpellier France, Vol. 2, pp. 767-173.

Tisseyre B., Mazzoni, C., Ardoin, N., Clipet, C., 2001, Yield and harvest quality measurement in precision viticulture - application for a selective vintage. In: Proceedings of 3rd ECPA, edited by S. Blackmore, G. Grenier, Agro Montpellier , France, Vol. 1, pp. 133-138.

Journel, A.G., Huijbregts, C.J., 1978, Mining Gesotatistics, Academic Press eds., London, UK, 601 p

Lecoutre, J.P., Tassi, P., 1987, Statistique Non Paramétrique et Robustesse, (Non parametric statistic and robustness) Economica eds., Paris., France, 455 p.

Whelan, B.M., M^{c} bratney, A.B., Minasny, B., 2001. VESPER – Spatial prediction software for precision agriculture. In: Proceedings of 3rd ECPA, edited by S. Blackmore, G. Grenier, Agro Montpellier , France, Vol. 1, pp. 139-145.