- **"New approach for large scale experimental research in agriculture and ecology "**

Jiftah Ben-Asher[1], Amir Musari[1], Uzi Kafkafi[2] and Jorge de Jesus[1]
1.The Institute for Desert Research Ben-Gurion University of the Negev. 2. Faculty of agriculture the Hebrew University of Jerusalem Israel. Special thanks to Don Nielsen from the university of California Davis USA.

## 1. Introduction

In the past century the majority of the phenomena's were normally studied on a time scale and rarely on a spatial scale. This was due to the elevated computational needs and heavy mathematics procedures needed for spatial analysis.
Nowadays the problems of elevated computational needs for this studies are no longer an obstacle, and so the development of GIS (Geographical Information Systems) during the 90's, and the application of the theory of regional variables (geostatistics) to other field beside mining, permitted the study of phenomena on a spatial/temporal scale and this methodology has proved to be able to answer to specific problems related to Agriculture (Panagopolous and Beltrão, 1999; Ben-Asher et al 1983)

## 2. Objectives

Suggest another methodology (avenue) of research for agriculture and ecology by:
1. Gathering a considerable amount of data of bimass production factors and yield from a field plot, in order to understand variations in yield.
2. To produce interpolation maps in a raster (pixels) structure
3. To use each pixle to verify if near-by pixel can be grouped in polygons such that they can be taken as virtual replication and each polygon can be taken as a virtual treatment which can be differentiated from other polygons by statistical procedure (ANOVA)
4. To correlate several plant production factors to improve the estimation of the kriging interpolation maps (reduction of standard deviation).
5. Determining minimal sampling need for maximal results

## 3. Analyse of Variance or ANOVA

In a general way, the analysis of variance (base of the statistics of Fisher 1951) presented in figure 1, ignores the geographic localization of the environmental element, and considers that the value of the samples are defined from the addition of the average, deviation and error.

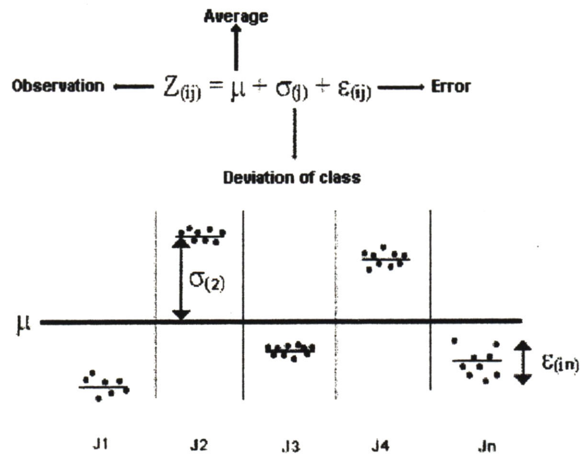



Figure 1 – Schematics of statistical elements of biomass production by various environmental structures. In analogy to analysis of variance (ANOVA), where J represents a treatment here j1 may represent the productivity of a forest, j2 may represent the productivity of an agricultural field etc.
The treatments are considered different or equal according to the standard-deviation and degree of freedom.
Despite being used for more than 50 years the ANOVA analysis has given time and time again proof to be of high utility, but it has permitted that some unnecessary constraints such as:

Imposition of unnecessary treatments
Selection of the places for sampling as "typical or representatives"
Usage of aleatory to avoid "bias"
Avoiding areas with high spatial variety
Assuming that variability is bad
Assuming that variability is an error
It doesn't thrust one sample
The results are not trustful for just one year and several years of experiments are required.
The design of causal blocks that is normally used in agricultural experiments it is used to answer only one question:
The treatments used caused a deviation ($\sigma^2$ on figure 1) from the average so big that the treatment can be considered different.
Despite the treatments used in experimentation the statistical drawing doesn't help the understanding of physical , chemical, or biological processes that occur on the field of specific plot
It doesn't identify the how one should gather the samples or the frequency
It doesn't identify where and the frequency of the sample gathering
It doesn't identify the size of the sample and the distance of the gathering
It doesn't identify the minimal area of the plot necessary for results.
And with respect to the objectives of the ICCAP projects it does not integrate spatial phenomena affected by global climate change.
To overcome the limitations of the traditional approach and the need to have better statistical methods for the research on global climate change there is a need to use a different and new statistical approach.

## 4. Analysis of regional variables – eostatistics

One of the important changes on experimentation was the introduction of the Analysis of the regional variables or more commonly known as Geostatistics. This new methodology is based on the principal that properties of one sample will be similar to its neighboring samples more than its similarity to a sample taken away from it. (Bresler,1989) and that each samples has also the information of its location (georeferenced) (Armstrong,1998)

## 5. Semi-variogram

On the more conventional statistics the data is mainly studied from the histogram graphic. In geostatistics the data gathered or used reflects a spatial phenomena, the histogram graphic is useless to analyze the spatial relations, so there is the need to used other graphic called the semi-variogram (Sousa et al. 1998). This graphic is based on equation 1 , the values of the semi-variogram are bigger as the distance of the sample increases, this distance is given by the module of the vector $\vec{h}$ that connects two nearby samples.

$$\gamma(h) = \frac{1}{2} N_{(h)} \sum [Z_i - Z_{i+}$$

Equation 1 –The semi-variogram determines how related in space are two or more samples, the most related the samples have lower value of $\gamma(h)$. The $N_{(h)}$ is the number of samples that can be grouped using vector h, the $Z_i$ represent the value of a sample, $Z_{i+h}$ is the value another sample that is distanced ||h|| from the initial sample ($Z_i$)

The spatial relation between the samples can be represented graphically by a graph denominated as semi-variogram, with the yy axis the value of equation 1 and on the xx axis the value of the module of vector $\vec{h}$.

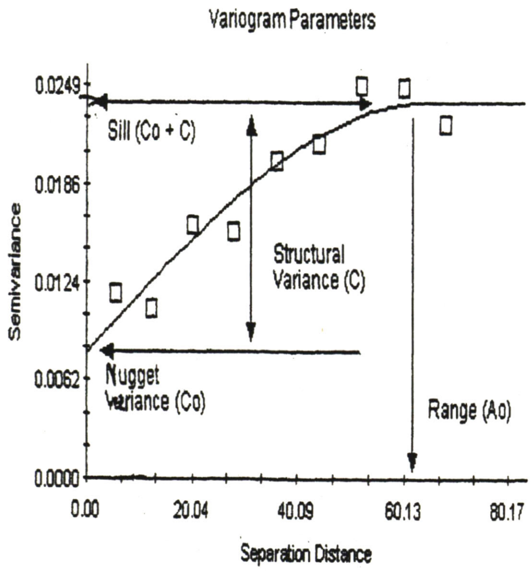
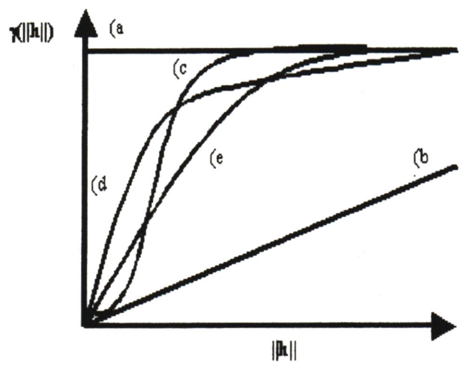



**Figure 2 – General display of semi-variogram and its its terminology, normally the semivariance value increases as the separation distance increases (module of vecto h), until it reacher the a distance which the samples are no longer related between them (called range Ao). The vale of the semi-variogram when reached the range is called Sill (Co+C), while the minimal value of semi-varience, or where the graph start is called the Nugget Efect or Nugget Variance (Co)**

From the semi-variogram (figure 2) it is possible to obtain information on interpolation, density of sampling, padrons of spacial distribution and to be able to performe spatial simulations. (Hartkmap et al., 1999).

## 6. Model of semi-variogram

The semi-varigram is a graphic of points, and there is the need for a continuous function that can descrive the behavior of the point graphic, or in other words, there is the need for a model. The problem to find a model that describes the point behavior, is that this model has to obey to certain mathematical rules that are hard to determine (Journel et al., 1978).. To overcome this mathematical problems, one should use only models have can fit this mathematical standart and that have given proof before that they work. The model picked is the one that give the best result using the least squares methods or other more advanced and modern techniques like the Cressie goodness of fit (Clark et al., 2000).




•

Figure 3 - Example of several possible models that can be used to descrive the behavior of the Semi-variogram. The nugget efect model a); Linear model b); Gaussian model c);Exponencial model d) and the Spherical model e). All this models use the range (Ao), sill (Co+C) and nugget value (Co) as parameters. Source Geographical Information Systems for Hydrology

After the determination which of the models that best describes the behavior of the semi-variogram, the algebra equation of it will be used for all the necessary calculations and estimations.

## 7. Interpolation procedure

Interpolation is the process by which an unknown value is obtained from other values that are nearby and their values are known (Clark, 1979). The methods of estimation in geostatistics consider that the unknown value is obtain by an average podenration, using the other known values. The way that the ponderation values are determined varies from each method

$$z(x_0)^* = \sum_{i=1}^{n} w_i . z(x_i) \quad \wedge \quad \sum_{i=1}^{n} w_i = 1$$

**Equation 2 – The unknown value $z(x_0)$ is obtained by the average ponderation of the known values. The poderations are given by $w_1$ and $z(x_i)$ are the known values. The sum of all the ponderation factors have to be identical to 1**

The ponderation values can be determined by using statistical methods (stoachistcs) or by non-statistical methods (deterministics), of course the stoastichics method requiere more calculations and computer power.

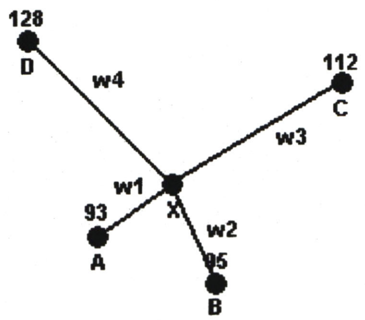



**Fig 4 – The value X is obtained from the ponderation of the values of A,B,C,D . As explained before, in geostatics there is the assumption that the value of X will be more similar to A and B than C and D due to the closest proximity of the first. The ponderations are given according the distances and the value differe according the method of estimation used.**

## 8. Ordinary Kriging

Ordinary Kriging or simplely Kriging is based on the acronym B.L.U.E , that is, "Best Linear Unbiased Estimator" (Shibli, 1997). This is a stochastic method that estimates the ponderations based on the assumption that the average of estimation error has to be zero and that the variance of the error has to be minimal.

$$R(x_o) = z(x_o)^* - z(x_o)$$

**Equation 3 - The error $R(x_o)$ is simplely the diference between the value estimated $z(x_o)^*$ and its real values $z(x_o)$**

This basic assumptions are not as simple has they look like, because they require the usage o matrix with a very high number of unknowns,and therefore a high computational needs. If one jumps some mathematical explanations, it reaches the stage of the kriging process is write has the matrix system below.

$$\underset{(n+1)x(n+1)}{\overset{C}{\begin{bmatrix} C_{11} & \cdots & C_{1n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ C_{n1} & \cdots & C_{nn} & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix}}} \cdot \underset{(n+1)x1}{\overset{w}{\begin{bmatrix} w_1 \\ \vdots \\ w_n \\ \mu \end{bmatrix}}} = \underset{(n+1)x1}{\overset{D}{\begin{bmatrix} C_{10} \\ \vdots \\ C_{n0} \\ 1 \end{bmatrix}}}$$

System 1 – Notation of the kriging method in matrixs (Armstrong,M. 1998)

The matrix C contains the co-variances between the samples that are around the sample to be determined. The matrix w contains the ponderations and a parameter called Lagrange parameter, and the matrix D contains the co-variance between the sample to be determined and the sorrounding ones. From the system of matrix, it is easy to notice that this method can determine the relations between the known and the unknown values

One important characteristic of Kriging is that aside from interpolating the necessary values, it also determines the standard deviation of the calculated value, (Clark, and Harper 2000).

The standard deviation of the interpolated value is given by:

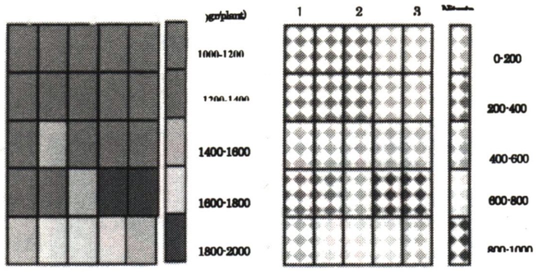
$$\sigma_r^2 = \sigma^2 - w.D \qquad [4]$$

Equation4 calculates the standard deviation on the interpolated value ($\sigma_r^2$) based on the standard deviation of the samples , and the matrix with the values of the ponderations (w) and the co-variance values contained in matrix D.

## 9. Geographical Information Systems (GIS)

GIS can be divided in 4 sub-systems 1) A sub-system that introduces and gather data and that processes spatial information from several sources like scanned maps, satellites images, databases etc.2) Storage of data and organization of spatial information, so that they can be easily recovered, edited and updated 3) A sub-system that manipulates, and analises data in such a way that it can estimate parameters, or performe modulation functions 4) Output of data in maps, graphics or databases. Thus, GIS is basically a tool that permits the production of maps, analyze spatial data, produce more data (results) from initial data sets and hence when combined with kriging may produce some virtual replications with standard deviation to enable the calculations of statistical significance of the data. This feature is demonstrated in figure 5.

Fig 5. Molti-Layer Analysis (Arcview GIS)




The squares in the right side of fig.5 represent several levels of measured soil nitrate concentration and are considered the basic layer. The squares on the left side of fig. 4 are the various levels of yield. This is the second layer. It can be seen that to each level of nitrate we identified a level of yield. Moreover the standard deviation of each group was also determined by the Arcview software .from which we could elaborate table 1

Table 1 Schematic presentation of Arcview output

| **$NO_3$ level** $gm^{-3}$ | **No of pixels** | **Average yield** g/plant | **Standard deviation** | **Significance level, t-test** |
|---|---|---|---|---|
| **X** | **Nx** | **$a_x$** | **Sx** | $\frac{a_x - a_y}{\sigma_x + \sigma_y} \geq t$ |
| **Y** | **Ny** | **$a_x$** | **Sy** | |

In table 1 X and Y are two virtual treatments that emerged from the interpolation of the measured data., each one of them was is characterized by N number of pixels average yield and standad deviation ( column no.4 ) such that the significance level can be calculated fron the equation in column no.5. Thus the measured results were used to generate semivariograms and kiging from which the numerical values were analyzed by Arcview software.

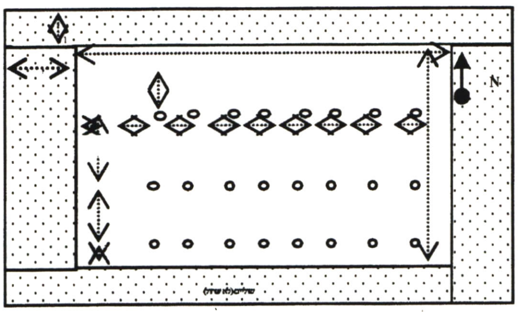
## 10. Materials and methods

- The crop: Cherry Tomato *Lycopersicom .escalentum var Mill Cherry species)*
- The research environment was commercial fields of greenhouses in the southern Mediterranean coast of Israel. The geostatisical analysis was made on a large number of parameters that was measured within two types of growing conditions. Conventional an organic agriculture. We compared conventional and organic soil environment
- Vegetative growth parameters
- Physiological dynamic parameters

Yield and its quality

The sampling plan is shown in Fig.5.

Fig 5. the sampling point in the 1000m$^2$ greenhouse.



Some of the instruments that were used in the research are shown below. In general we established a mobile research lab the can be used on a large area within a short sampling time.

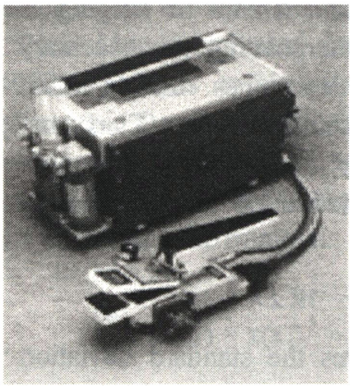


Fig 6 the portable IRGA

The IRGA ( Infrared gas analyzer ) was used to measure the following parameters on an instantaneous basis from a single leaf. : Photosynthesis and ,transpiration rate at noon time stomata resistance VPD leaf temperature

Soil parameters included soil respiration with the respiration chamber shown below

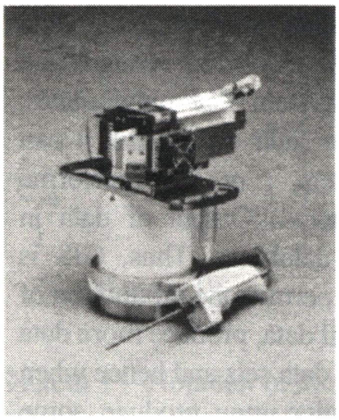


Fig 7 Soil respiration chmber

soil temperature relative humidity of the green house soil water content with TDR soil salinity with TDR etc. A special environmental integrator (Fig 8) was devices in order to measure the above variables.

Fig.8 The mobile environmental integrator

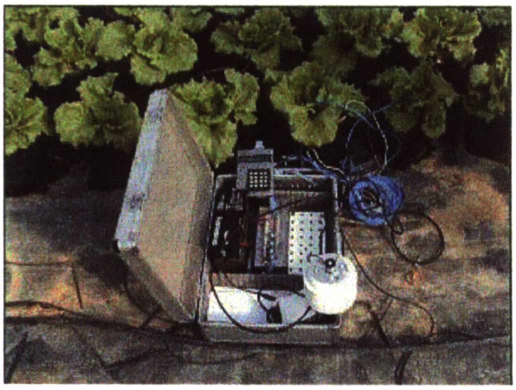
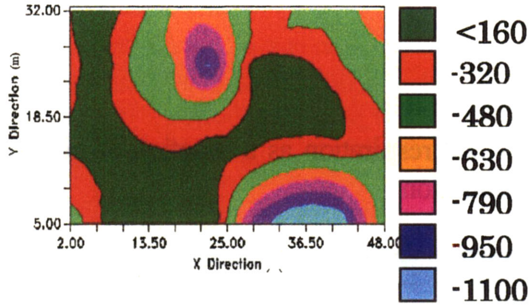


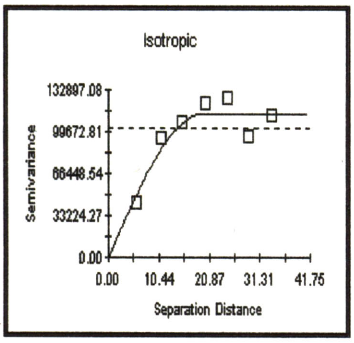
**Fig 9b. Kriging of nitrate distribution in a commercial greenhouse**




## 11. Results and discussion

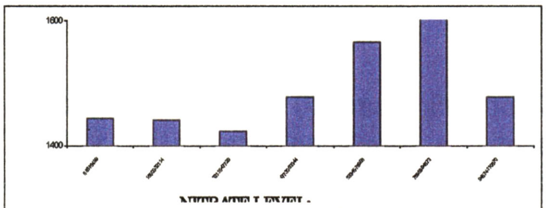
The distribution of nitrate concentrations in the field is given in Fig.9 a and b

Fig9a) The semivariogram and b) the resulting distribution of nitrate concentrations




From Fig 9b it can be seen that nitrate distribution in the field was highly heterogeneous and the response of the yield at several sites in the field was strongly affected by this heterogeneity. This can be seen from Fig. 9 in which we assigned the yield to its site and nitrate concentration at this site.

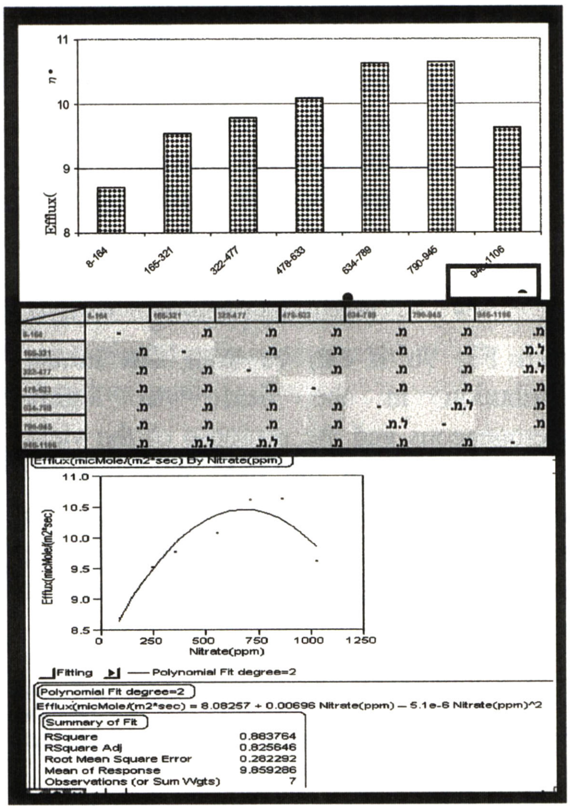
Fig 10 Tomato yield as a function of nitrate concentration




| | 8.86-164.99 | 165.00-321.14 | 321.15-477.29 | 477.30-633.44 | 633.45-789.58 | 789.59-945.73 | 945.74-1106.70 |
|---|---|---|---|---|---|---|---|
| 8.86-164.99 | - | NS | S | S | S | S | NS |
| 165.00-321.14 | NS | | S | S | S | S | NS |
| 321.15-477.29 | S | S | - | S | S | S | S |
| 477.30-633.44 | S | S | S | - | S | S | NS |
| 633.45-789.58 | S | S | S | S | - | S | S |
| 789.59-945.73 | S | S | S | S | S | - | S |
| 945.74-1106.70 | NS | NS | S | NS | S | S | - |

The measurement correspond to the values in the scale of the kriging in fig 9b. The lower part of Fig 10 is a table of significance which was obtained form the GIS analysis. It generally suggests that the differences between most virtual treatments were significant.

Fig11 soil respiration as a function of nitrate concentration




Comparing figures 10 and 11 one can see that there is a correlation between the yield of tomato and soil respiration. Until a concentration of 950 ppm nitrate root respiration and productivity increased. At a higher nitrate concentration both respiration and yield started to decrease and results were significant.

## 12. Concluding remarks

We gave here an example of suggested technology for data collection and analysis in the ICCAP project. Though its first test was made on a small scale the following conclusions apply also for the large scale

1. The use of geostatistics combined with GIS may help integrating the various sources of data in the project.
2. It helps analysis of a large number of affected and effecting variables and selecting the dominant factor affected by GCC.
3. The data are based only on spatial distribution and only very little involved temporal effects. This should be included in the ICCAP project in order to combine empirical observations with deterministic equations.

## 13. References


- Ben-Asher,J. Warrick,A. Matthias,A. 1983. Bare-soil evaporation determined *in situ* by infrared thermometry. *J.Hydrology* 69:325-334
- Bresler,E. 1989. Estimation of Statistical Momments of Spatial Field Averages for Soil Properties and Crop Yields. Soil Sci. Soc. Am. J. 53:1645-1643
- Clark, Isobel, *Practical Geostatistics*, 1979, Applied Science Publishers Ltd
- Clark,I.; Harper, B. ; 2000. Practical Geostatitcs. Geostokos Editions, London
- Fisher R.A. The design of experiments 6th edition Oliver and Boyd 1951
- Hartkamp,A. De Beurs,K.;Sein,A.;White,J. 2001. Interpolation Techniques for Climate Variables
- Journel, A.G. and C.J. Huijbregts. 1978. Mining Geostatistics. Academic Press, New York
- McBratney, A.B., Webster, R., Burgess, T.M., 1981. The design of optimal sampling schemes for local estimation and mapping of regionalised variables. Computers and Geosciences 7, 331–334.
- Panagopoulos,T. Jesus,J. Beltrão,J. 2001. Soil geostatistical techniques applied to

sprinkle irrigation experiments. 7th International meeting of Soils with Meditterranean type of climate, (ed.)