

APPLICATION OF QUALITATIVE ANALYSIS TECHNIQUES TO THE ENVIRONMENTAL MODELING OF PLANT SPECIES CULTIVATION

I.K. KALAVROUZIOS^{1,*}
V.A. VISSIKIRSKY²
V.S. STEPASHKO²
P.H. KOUKOULAKIS³

¹*Department of Environmental and Natural Resources Management, University of Ioannina, Greece*

²*Department for Information Technologies of Inductive Modeling, The International Center of Information Technologies and Systems of the National Academy of Sciences of Ukraine
40 Glushkova ave., Kiev, Ukraine*

³*Soil Science Institute, Nagref, Atfalidos 8, GR-544 54, Thessaloniki, Greece*

Received: 02/10/09
Accepted: 10/06/10

*to whom all correspondence should be addressed:
e-mail: ikalavru@cc.uoi.gr

ABSTRACT

The main issue discussed in the paper is the application of qualitative analysis techniques when short data samples are available at the initial stage of experiments carried out to study behaviors of agricultural plant species under different environmental conditions. The proposed techniques are described on the examples of two tasks: analyze behaviors of chemical elements in plant species irrigated with treated wastewater (on the example of Broccoli cabbage), as well as in plant species cultivated near the highway (on the example of *Lycopersicum* species). The system of ecological interactions is decomposed into individual subsystems with sufficiently small number of variables describing concentrations of elements. The Group Method of Data Handling (GMDH) is applied to obtain the regressions between variables within subsystems. The Qualitative Description based on signs combinations of differences for a regression function is employed to qualitatively analyze the relationships among variables, both within individual subsystems and within the system in whole. This approach makes it possible to identify main behavior features of individual objects/relationships between objects, classify behaviors according to different criteria of similarity/distinction, provide feature matching, etc. The obtained results then can be used to design the next-stage experiments and construct more comprehensive models of plant species behaviors.

KEYWORDS: Environmental modeling, agriculture, wastewater, road pollution, GMDH, qualitative analysis.

1. INTRODUCTION

The aim of the present paper is to study the application of Qualitative Analysis techniques to the results of experiments carried out to study different aspects of environmental impact on growth of cultivated plant species. Small numbers of data samples are inherent in the considered field because plant growth is slow process, while a large number of input variables are required to be involved in the analysis. Experiments designed to identify species characteristics under different conditions are conducted by series, during long-term period. Therefore, it is important to apply qualitative analysis techniques, enabling us to make preliminary assessment of features of plant species behavior. The results of an initial qualitative analysis play the role of assumptions that can be used to design the next-stage experiments. A number of possibilities of structural and qualitative analysis have been reported (Vissikirsky *et al.*, 2005; 2007).

In order to demonstrate capabilities of the proposed techniques, two typical examples are described here: analysis of concentrations of chemical elements in plant species irrigated with treated municipal wastewater, on the example of Broccoli (*Brassica oleracea* var. *Italica*) (Kalavrouziotis *et*

al., 2007b; 2008), as well as in plant species cultivated near the highway, on the example of *Lycopersicum esculentum* (Kalavrouzios *et al.*, 2007a).

For the purpose of assessing of the environmental impact on cultivated plant species, various modeling solutions were presented by authors in other papers depending on the objective, system of ecological interactions, volume of measurement data, etc (Drakatos *et al.*, 2001; Vissikirsky *et al.*, 2008a; 2008b). In particular, the Group Method of Data Handling (GMDH) (Farlow, 1984) was applied to estimate different aspects of plant species behavior: impact of concentrations of chemical elements contained in wastewater on the mechanical properties of trees (Kalavrouzios *et al.*, 2002), height and mortality development in trees irrigated with wastewater (Kalavrouzios *et al.*, 2004), etc.

With incomplete knowledge about structure of an ecological system, and with small number of measured data, the qualitative analysis would allow the researcher to grasp main the features of system behavior, consider synergetic or antagonistic interactions between different variables in a suitable interpretation form, and make decisions about the direction of the next-stage experiments.

In addition, one of the problems of environmental modeling is the great number of system variables. The basic approach to resolve this problem consists in the decomposition of a modeled system into subsystems, as well as in modeling and analysis of individual subsystems with a number of variables consistent with the volume of measurement samples available. In this case, however, the techniques will be required to provide qualitative analysis of total system behavior. The "qualitative" approach would allow us to overcome the problem of great number of variables with small volume of data samples, and give a total though "shallow" picture, of the considered processes.

The analysis in this paper was carried out within the chain of interactions "soil - roots - leaves" (and "heads" in case of Broccoli). Different regression relationships within the chain were obtained with the GMDH. On the basis of so called Qualitative Description (QD) of a function (Vissikirsky, 2001), which is described below, qualitative relationships (e.g., "increase", "decrease") were derived from the regressions. Further qualitative analysis of chemical elements behavior in soil and in different parts of plants was carried out over the set of QDs obtained for system variables.

2. QUALITATIVE DESCRIPTION OF A FUNCTION BASED ON COMBINATION OF SIGNS OF DIFFERENCES

Knowledge about relationships among qualitative features of variables (such as low or high, increase or sharp increase) can be called qualitative knowledge (Kuipers, 1986). Qualitative relationships can be interpreted as ordinal relations, which are described in (De Kleer, Brown, 1984; Kuipers, 1984). In the qualitative simulation systems, the qualitative relationships are applied to the set of landmark values reflecting a change in behavior of structural components. Qualitative simulation is intended for the derivation of possible qualitative behaviors of a model, given certain set of qualitative constraints imposed on its structure.

The considered task consists in finding qualitative relationships in numerical data. One of the examples of "qualitative data mining" is described in (Bratko, Suc, 2003), where so the called *qualitative decision trees* are obtained on the basis of *qualitatively constrained functions* (QCFs). For example, the QCF $Z = M^{+-}(X, Y)$ says that Z monotonically increases in X and decreases in Y .

Here, we propose the Qualitative Description (QD), which makes it possible to find and analyze qualitative features in detail including non-monotonicity (extrema, concavity/convexity, inflection, etc.) within different measurement intervals. By Qualitative Description (QD) of a function it is understood a string of signs of differences obtained for discrete values of the function on a time interval. In general, the QD which is required to represent main behavior features involves differences of first three orders. Therefore, each QD includes three strings over the alphabet $\{+, -, =, *\}$. The symbols "+" and "-" are the signs of difference, "=" denotes zero value of difference, "*" means "do not care" similarly to that used in classifier systems (Booker *et al.*, 1989). Qualitative features of functions such as extrema and their mutual positions, monotonicity, degree of growth, etc., can be derived from combinations of QD signs. Such representation can be viewed as sufficient for initial analysis with lack of detailed information about system behavior. This way also enables us to assess individual subsystems and as well as system in whole.

The function $f(t)$ is investigated on finite closed time interval $[t_1, t_n]$. As a result of simulation or measurements, we obtain the values $x_i = f(t_i)$ for finite number $i = \overline{1, n}$ of discrete time-points t_i on $[t_1, t_n]$; a step h_t determines the intervals $[t_i, t_{i+1}]$.

Let

$$\begin{aligned}\Delta_i^1 &= x_{i+1} - x_i, \\ \Delta_i^2 &= \Delta_{i+1}^1 - \Delta_i^1 = x_{i+2} - 2x_{i+1} + x_i, \\ \Delta_i^3 &= \Delta_{i+1}^2 - \Delta_i^2 = x_{i+3} - 3x_{i+2} + 3x_{i+1} - x_i,\end{aligned}\quad (1)$$

are descending differences approximating $f(t)$ on $[t_i, t_{i+3}]$. The Qualitative Description

$Q(f)$ of function $f(t)$ is a triple:

$$Q(f) = (q^1, q^2, q^3), \quad (2)$$

where

$$q^1 = \{s_i^1\}^{n-1}, q^2 = \{s_i^2\}^{n-2}, q^3 = \{s_i^3\}^{n-3}, \quad (3)$$

$$s_i^r = \begin{cases} +, & \text{if } \Delta_i^r > 0, \\ =, & \text{if } \Delta_i^r = 0, \\ -, & \text{if } \Delta_i^r < 0, \\ *, & \text{don't care.} \end{cases} \quad (4)$$

Here, $\{s\}^k$ denotes a string of length k of symbols s over the alphabet $\{+, =, -, *\}$. Each component q^r ($r=1,2,3$) describes the sequence of signs in differences of order r , in which the symbol "*" may be included. The Qualitative Description can be represented different forms.

The QD with three order differences ($r=3$) is practically sufficient to describe all main features of function variation. In this case, minimum number of discrete values of function must be no less than four. It is also possible to use a reduced description with $r=2$ or even $r=1$, with lesser minimum number of discrete values required.

In case if there is a sequence of the same symbols of a (sub)string then it can be represented as ks_i^r , where k is the number of alike symbols. For example, $(++++***)$ is equal to $(3+2*1-)$ or $(3+2*-)$. If for all q^r their strings consist of equal signs then QD can be written in a compact form, for instance, $Q = (+, -, =)$.

Let us assign to the four of values $x_i, x_{i+1}, x_{i+2}, x_{i+3}$ the set of differences whose expressions include those values. Let such minimal fragment of QD be called i -th qualitative cell c_i if:

$$c_i = (c_i^1, c_i^2, c_i^3) = (s_i^1 s_{i+1}^1 s_{i+2}^1, s_i^2 s_{i+1}^2, s_i^3), \quad (5)$$

The qualitative evolution operator of a function $f(t)$ is the totally ordered sequence of its qualitative cells $c_i, i-2=1, n-2$. This notion is close to the operator of evolution $t \Rightarrow f(t)$ utilized in ordinary differential equations to describe movement of phase point along integral curve $f(t)$. This notion is useful for the analysis of function features when moving along time axis.

Derivation of qualitative features from Qualitative Description

The most widely used qualitative features arising in the movement along time axis are: increase-decrease, inflections, sequence and mutual positions of extrema, rates of growth, forms of peaks. Prototypes of typical features are given below. Here, the symbol "/" is used to enumerate optional signs combinations.

$$\begin{aligned}s_i^1 &= (+/-) && - \text{monotonicity (increase/decrease),} \\ (s_i^1, s_{i+1}^1) &= (+-/-+) && - \text{maximum /minimum,} \\ s_i^2 &= (+/-) && - \text{concavity/convexity,}\end{aligned}$$

$$(s_i^2, s_{i+1}^2) = (+ - / - +) \quad - \text{ inflection (outward/inward).}$$

When analyzing the qualitative evolution operator or deriving more complicated features, it is required to take into account the constraints on signs combinations of adjacent cells.

For example, the admissible c^2 and c^3 values for $c^1 = (+++)$ are:

$$(c^2, c^3) = (++,+ / -), (+ =, -), (+ -, -), (---, - / +), (- =, +), (- +, +), (= +, +), (==, =) \quad (6)$$

In case of nonmonotonic behavior, $c^1 = (++-)$, the c^2 characterizes a form of growth of function right before its maximum, while for $c^1 = (+--)$, the c^2 does a form of fall after reaching its maximum. In case of $c^1 = (+--)$ and $c^2 = (---)$, the transition from $c^3 = (-)$ to $c^3 = (+)$ reflects a tendency to have more sharp maximum and fall.

In order to describe the behavior for both sides of maximum, it is necessary to obtain a fragment of a qualitative evolution operator $g_{i,i+1}$ by merging the adjacent cells $(c_i) = (+ + -, *, *)$ and

$$(c_{i+1}) = (+ --, -, *):$$

$$g_{i,i+1} = (c_i, c_{i+1}) = (+ + --, * - *, **) \quad (7)$$

The cases of behavior are obtained when replacing "*" by signs admissible for a given position.

The Qualitative Description is a qualitative analysis tool dealing with a set of characteristic features of a function (model). Therefore, the experimental measurements shall be preliminarily statistically processed to obtain representative features relevant to the stage of experiments. Here, the QD was applied to the GMDH-models that estimate subsystems.

Let us now consider two examples using the GMDH and Qualitative Description, to assess behavior of chemical substances in soil and in plant species under different conditions.

3. EXAMPLE 1: ELEMENTAL FLUCTUATIONS IN SOIL, ROOTS, LEAVES, AND HEADS OF BROCCOLI PLANT SPECIES, AS AFFECTED BY DIFFERENT WATER TREATMENTS

The experiments were conducted in a greenhouse of the Department of Environmental and Natural Resources Management at the Ioannina University, located in Agrinion, Greece, where the effects of repeated applications of municipal reclaimed wastewater, compared to the fresh irrigation (control) water, were studied on soil and different parts of *Broccoli* (*Brassica oleracea var. Italica*) plant species (Kalavrouzotis *et al.*, 2007b; 2008).

The experiments were carried with the following design: two water sources (wastewater and control water) in six replications, and two Brassica species (Brussels sprouts and Broccoli) with a total of (2 species) x (2 treatments) x (6 replications) = 24 experimental plots, of 2.5x1.8m² size. The irrigation time was determined on the basis of 40-45% water depletion of the soil field capacity, and the application of wastewater and control was done by means of a hand driven watering can. Soil samples from the depth of 0-30cm were taken (a) short before the commencement of the experiment (b) 8 weeks after planting, and (c) 16 weeks after planting (just before harvesting). Also, samples of leaves and roots were taken in 8th and 16th weeks after planting. Harvesting was done in the 16th week, and the experiments were completed.

The average measured concentrations of chemical elements in water, soil, and parts of plant species are given in Tables 1, 2, and 3 respectively. Here, CW and WW refer to the control water and waste water, and 1CW - 3CW and 1WW - 3WW to the sampling and measurement period, conducted at 0th, 8th, and 16th week after planting, respectively, for control water and wastewater, respectively.

Table 1. Average elemental content of control (CW) and treated municipal wastewater (WW)

| | P Mg l⁻¹ | K Mg l⁻¹ | Ca mg l⁻¹ | Mg Mg l⁻¹ | Mn Mg l⁻¹ | Zn Mg l⁻¹ | Fe Mg l⁻¹ | B Mg l⁻¹ | Cu Mg l⁻¹ |
|----|--------------------------------|--------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|--------------------------------|---------------------------------|
| CW | 0.05 | 0.88 | 49.03 | 4.2 | 4.11 | 6.22 | 35.56 | 0.67 | 1.78 |
| WW | 0.64 | 16.14 | 90.74 | 21.63 | 84.54 | 109.76 | 102.89 | 1.18 | 2.73 |

Table 2. Average elemental contents of soil

| Soil | 1CW | 2CW | 3CW | 1WW | 2WW | 3WW |
|---------------------------------|------------|------------|------------|------------|------------|------------|
| P mg kg⁻¹ | 18.86 | 24.62 | 46.20 | 14.32 | 30.54 | 41.16 |
| K mg kg⁻¹ | 121.7 | 121.7 | 113.3 | 131.7 | 130.0 | 118.0 |
| Ca meq 100g⁻¹ | 1.26 | 1.39 | 1.26 | 1.27 | 1.37 | 1.04 |
| Mg meq 100g⁻¹ | 0.27 | 0.36 | 0.35 | 0.29 | 0.30 | 0.45 |
| Mn mg kg⁻¹ | 15.27 | 20.55 | 15.37 | 14.99 | 20.77 | 15.36 |
| Zn mg kg⁻¹ | 3.61 | 4.83 | 5.38 | 3.73 | 4.57 | 4.64 |
| Fe mg kg⁻¹ | 18.22 | 19.94 | 16.87 | 14.81 | 17.61 | 13.99 |
| B mg kg⁻¹ | 0.34 | 0.43 | 0.21 | 0.34 | 0.44 | 0.27 |
| Cu mg kg⁻¹ | 1.73 | 1.6 | 1.37 | 1.47 | 1.57 | 1.30 |

According to the t-test conducted on the elemental composition of soil, the differences found between sampling periods as well as between the treatments applied, i.e. CW and WW, were generally non significant. On the other hand, in the case of plant analyses data, there were some statistically significant differences, such as in the case of Fe and Mn (Table 3).

Table 3. Average chemical contents in roots, leaves, and heads

| Objects | Roots | | | | Leaves | | | | Heads | |
|-----------------------------|--------------|------------|------------|------------|---------------|------------|------------|------------|--------------|------------|
| | 2CW | 3CW | 2WW | 3WW | 2CW | 3CW | 2WW | 3WW | 3CW | 3WW |
| P% | 0.23 | 0.19 | 0.25 | 0.17 | 0.41 | 0.20 | 0.39 | 0.19 | 0.425 | 0.427 |
| K% | 2.64 | 2.73 | 2.60 | 1.26 | 5.17 | 2.59 | 5.13 | 2.88 | 3.23 | 3.523 |
| Ca% | 0.86 | 1.05 | 0.77 | 1.26 | 3.66 | 6.16 | 3.38 | 5.74 | 0.941 | 1.093 |
| Mg% | 0.60 | 0.48 | 0.54 | 0.39 | 0.62 | 0.65 | 0.63 | 0.53 | 0.352 | 0.412 |
| Mn µg g⁻¹ | 29.03 | 47.91 | 25.78 | 94.7 | 49.73 | 51.09 | 48.12 | 63.03 | 26.70 | 32.42 |
| Zn µg g⁻¹ | 33.23 | 24.37 | 31.35 | 27.05 | 32.56 | 15.92 | 34.15 | 18.05 | 36.65 | 38.62 |
| Fe µg g⁻¹ | 379.5 | 758.7 | 432.85 | 1022.0 | 90.58 | 85.65 | 93.4 | 91.97 | 60.68 | 61.80 |
| B µg g⁻¹ | 31.37 | 21.74 | 29.62 | 52.59 | 25.30 | 21.19 | 24.63 | 21.44 | 18.92 | 21.04 |
| Cu µg g⁻¹ | 8.73 | 9.20 | 7.75 | 8.42 | 12.63 | 7.30 | 7.65 | 7.30 | 8.15 | 7.16 |

The objective of this example is to identify and analyze behavior of chemical elements concentrations in soil, as well as in roots, leaves, and heads of the Broccoli plant species. Here, the studied system involves four objects: soil (S), roots (R), leaves (L), and heads (H), and has the following structure:

$$W(E) \rightarrow S(E) \rightarrow R(E) \rightarrow (L(E) \rightarrow H(E)) \quad (8)$$

where $W = \{CW, WW\}$ is water cases: CW - control water, WW - wastewater; $E = \{P, K, Ca, Mg, Mn, Zn, Fe, B, Cu\}$ are concentrations of chemical elements.

In this structure, a contiguous relationship, e.g. $S \rightarrow R$, is the direct factor for which it is assumed that S is the main object effecting on R; S is the indirect factor for L. Water content is assumed to be a basic factor effecting on the whole chain via soil.

Regression models for individual chemical elements

Due to short samples at this stage, the design of regression models will be limited to the direct factors for each individual element. Because water content is approximately constant over the

survey period for the control and waste water, the models are be designed for each water irrigation case. Here, the regression models are obtained with the Multilayered Iterative Algorithm (MIA) based on the GMDH method (Kalavrouzios *et al.*, 2002; Farlow, 1984). They are also used to get Qualitative Descriptions and carry out qualitative analysis of similarity or distinction in behavior of different chemical elements/objects.

Let us estimate direct factors at final period: $S3 \rightarrow R3$, $R3 \rightarrow L3$, $L3 \rightarrow H3$, and $R3 \rightarrow H3$, where $S3$, $R3$, $L3$, and $H3$ are concentrations in soil, roots, leaves, and heads at the 16th week respectively.

Let us take the P element as an example. Consider $S3 \rightarrow R3$ for the case CW :

$$P_R = 0.68914 - 0.026786 * P_S + 0.00033112 * P_S^2 \quad (9)$$

Fig. 1 shows the plots of measured and estimated results.

The regression model for $R3 \rightarrow L3$, CW has the following form:

$$P_L = 0.21085 - 0.70378 * P_R + 2.9251 * P_R^2 \quad (10)$$

The regression model for $L3 \rightarrow H3$, CW has the following form:

$$P_H = -0.76683 + 10.882 * P_L - 23.699 * P_L^2 \quad (11)$$

The regression model for $R3 \rightarrow H3$, CW has the following form:

$$P_H = 0.089037 + 3.3073 * P_R - 7.5226 * P_R^2 \quad (12)$$

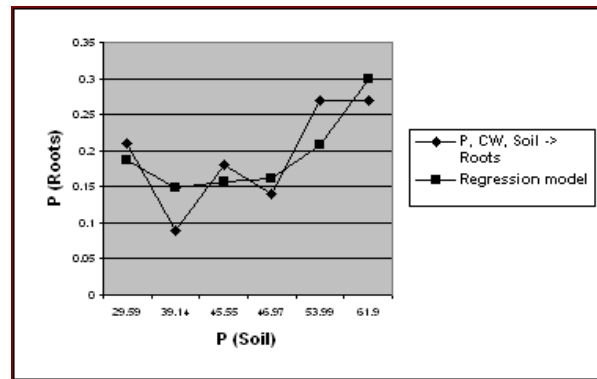


Figure 1. Measurements and regression of $S3 \rightarrow R3$ for P , under the effect of CW

Analysis based on Qualitative Descriptions

We can derive the differences between adjacent values proceeding from the estimated values over the time intervals. The obtained Qualitative Descriptions for different relationships and water irrigation conditions are shown in Table 4. It is also reasonable to include the descriptions for the difference ($WW-CW$) in order to analyze distinctions between the two water irrigation modes. In addition, this makes it possible to involve in the analysis the heads of Broccoli, as they were measured once at the final stage only.

Since there are three measurements for soil, and two measurements for roots and leaves, the QD structure will be reduced according to the number of available values. We can obtain two the first-order differences and one second-order difference for soil, and only one the first-order difference for roots and leaves. Therefore, the cell $c_i = (c_i^1, c_i^2) = (s_i^1 s_{i+1}^1, s_i^2)$ can be used for soil, and the first-order difference $c_i = (c_i^1) = (s_i^1)$ for the other objects.

In order to facilitate interpretation of the results, Table 5 shows the forms of curves for different combinations of signs. Here, the signs "=" for the second-order differences are not included.

Let us first consider the chain of factors for individual chemical element, e.g. P (Fig. 2).

For (CW): $Soil$ ($++$, $+$) means that P is monotonically accumulated in soil with acceleration, where "monotonically" is described by $c^1 = (++)$, while "with acceleration" by $c^2 = (+)$.

For (WW): $Soil$ ($++$, $-$) means that P is monotonically accumulated in soil with deceleration, where "with deceleration" is described by $c^2 = (-)$.

For (WW-CW): Soil (+,-,-) means the change in difference between **P** accumulation in soil when applying WW and CW. At the first stage, accumulation with WW was faster (the first plus), and with CW at the second stage (the second minus).

The **P** concentration in roots falls for both CW and WW, but faster for WW. "Faster" is derived from "-" for (WW-CW).

The **P** concentration in leaves falls for both CW and WW, but more slowly for WW. "More slowly" is derived from "+" for (WW-CW).

The Qualitative Description provides a simple classification of elements or objects with common features of interest. For this purpose, the user can write a request string in the form of a QD describing those features. The result of request execution is the subset of elements (objects) that have the features described in the request (if any). The features of interest are described by signs combinations, while the remaining positions are marked with "*".

Table 4. Qualitative Descriptions for different water irrigation conditions

| | CW | | | WW | | | WW-CW | | | |
|-----------|-------|-------|--------|-------|-------|--------|-------|-------|--------|-------|
| | Soil | Roots | Leaves | Soil | Roots | Leaves | Soil | Roots | Leaves | Heads |
| P | ++,+ | - | - | ++,- | - | - | +,-,- | - | + | + |
| K | =-,- | + | - | --,- | - | - | --,- | - | + | + |
| Ca | +,-,- | + | + | +,-,- | + | + | --,- | + | - | + |
| Mg | +,-,- | - | + | ++,+ | - | - | -+,+ | - | - | + |
| Mn | +,-,- | + | + | +,-,- | + | + | +,-,- | + | + | + |
| Zn | ++,- | - | - | ++,- | = | - | --,- | + | + | + |
| Fe | +,-,- | + | - | +,-,- | + | - | +,-,- | + | + | + |
| B | +,-,- | - | - | +,-,- | + | - | ++,+ | + | + | + |
| Cu | --,- | + | - | +,-,- | + | - | +,-,- | + | + | - |

Table 5. Combinations of signs and relevant forms of curves

| | | | | | | | |
|------|------|-------|-------|------|------|------|------|
| ++,+ | ++,- | +=-,- | +,-,- | -+,+ | =-,+ | --,+ | --,- |
| | | | | | | | |

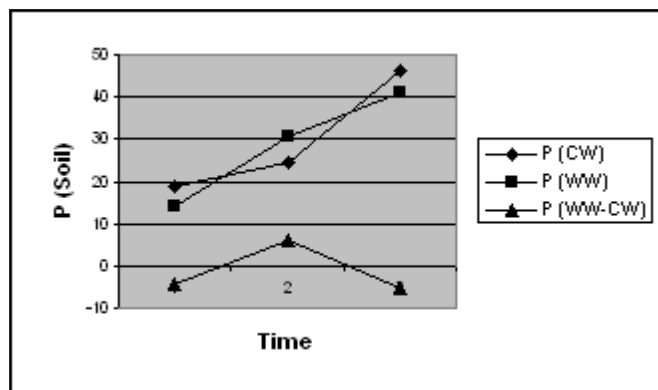


Figure 2. Behavior of **P** in soil

For example, the request:

$$\text{For } (W): \{\text{Soil } (++, *)\} \Rightarrow \{P, Zn\} \tag{13}$$

where $W = WW \& CW$, is directed to obtain the subset of elements with monotonic increase of concentrations in soil, with no care of how fast that increase occurs. The result is the subset of elements $\{P, Zn\}$.

In case to clear up what elements in soil with *WW* (if any) are increasingly accumulated, and are monotonically greater in comparison with *CW*, the following request can be formulated (see Fig. 3):

$$\text{For } (WW-CW): \{\text{Soil } (++,*)\} \Rightarrow \{B\} \quad (14)$$

At the same time, $\text{For } (WW \text{ or } CW): \{\text{Soil } B (++,*)\} = \phi$, that is neither *WW* nor *CW* have monotonic increase in concentration of *B* (at the first stage only).

Other examples of requests are:

1) Monotonic increase of concentrations in all objects in *CW*:

$$\text{For } (CW): \text{Soil } (++,*) \& \text{Roots } (+) \& \text{Leaves } (+) = \phi \quad (15)$$

2) Monotonic increase of concentrations in roots and leaves:

$$\text{For } (W): \{\text{Roots } (+) \& \text{Leaves } (+)\} \Rightarrow \{Ca, Mn\} \quad (16)$$

3) Monotonic excess of concentration in roots, leaves, and heads with *WW* in comparison with *CW*:

$$\text{For } (WW-CW): \{\text{Roots } (+) \& \text{Leaves } (+) \& \text{Heads } (+)\} \Rightarrow \{Mn, Zn, Fe, B\} = \{A1\} \quad (17)$$

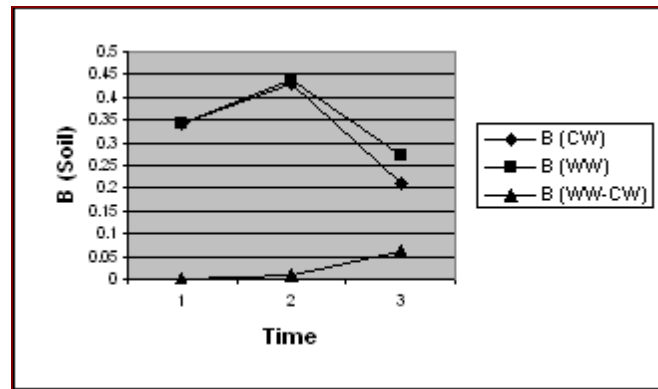


Figure 3. Behavior of *B* in soil

It is also possible to carry out a sequence of requests, by applying subsequent request to a subset obtained from previous request. The examples of those requests for the subset $\{A1\}$ (see (17)) are:

$$\text{For } (W): \{\text{Roots } (+) \& \text{Leaves } (+)\} \Rightarrow \{Mn\} \quad (18)$$

$$\text{For } (W): \{\text{Roots } (+)\} \Rightarrow \{Mn, Fe\} \quad (19)$$

$$\text{For } (W): \{\text{Soil } (++,*) \& \text{Roots } (+)\} = \phi \quad (20)$$

In addition, we can analyze similarity and distinction in behavior of chemical elements within the same object. The result (13) is the example of similar behavior.

In order to find distinction between *CW* and *WW* for elements in soil, the following request can be done:

$$\text{For } (CW): \text{Soil } (++,*) \& \text{For } (WW): \text{Soil } (-,*) \Rightarrow \phi \quad (21)$$

which means that there is no substantial difference in the behavior of elements, but by stages and intensity of concentration increase/decrease only.

With at least one change in corresponding positions of signs between *CW* and *WW*, the following pair of elements is obtained for soil:

$$\begin{aligned} CW: \text{Soil } Mg (+,-) \& WW: \text{Soil } Mg (++,) \\ CW: \text{Soil } Cu (-,-) \& WW: \text{Soil } Cu (+,-) \end{aligned} \quad (22)$$

It can be seen from Table 4 that decrease of concentration in roots and leaves can be observed for *P* (*W*), as well as for *Zn*, *B* (*CW*) and *K*, *Mg* (*WW*).

4. EXAMPLE 2: EFFECTS OF AIRBORNE CHEMICAL ELEMENTS ON SOIL, ROOTS, AND LEAVES OF LYCOPERSICUM PLANT SPECIES DEPENDING ON DISTANCE TO THE HIGHWAY

This example considers behavior of chemical elements concentrations in soil, leaves, and roots of *Lycopersicum esculentum* L. species cultivated near the main highway in the Lamias district, Greece (Kalavrouziotis *et al.*, 2007a). One transect at right angles away from the road was selected for this initial study. Samples of leaves and roots were collected at distances of 0, 5, 10, 20, 50, and 100 m from the road. Two soil samples were collected at each of the same locations (one at each of the depths of 0-5 and 5-15 cm).

The objective of this example is similar to the previous one, but here we shall analyze behavior of chemical elements depending on distance from the road. The studied system involves three objects: soil (S), roots (R), and leaves (L) with the following structure:

$$S(E, D) \rightarrow R(E, D) \rightarrow L(E, D) \quad (23)$$

where D is the distance from the road; $E = \{Pb, Al, Cd, Zn, Cu, Ni, Co, Fe, Mn\}$ are concentrations of chemical elements, which were measured in soil, roots, and leaves at distances of 0, 5, 10, 20, 50, and 100 m from the road.

Here, contiguous relationships $S \rightarrow R$ and $R \rightarrow L$ are the direct factors. In real situation there may exist other factors unaccounted in (23), for example, air environment, rainfall, etc.

The regression models analyzed here have the following form:

$$D \rightarrow E_i(C_j) \quad (24)$$

where $C_j \in \{S, R, L\}$, $E_i \in E$, i.e. change in concentration of each chemical element in soil, roots, and leaves depending on distance.

Tables 6 and 7 summarize the data of concentrations measured for soil, roots, and leaves, depending on distance and soil depth. It is seen in Table 7 that the concentration of Al in the roots is very high in comparison to the low Al level in the soil (Table 6). These differences are most probably due to the fact that most of the Al taken up by the plants is usually accumulating in the root system and only 10% of the total uptaken Al is transported to the above ground plant parts (Srivastava and Gupta, 1996).

The MIA regression models applied for individual elements are summarized in Table 8.

Table 6. Average elemental content (mg kg⁻¹) in soil adjacent to the national highway of Lamias – Athens, Greece

| | Soil depth (cm) | Distance (m) | Pb | Al | Cd | Zn | Cu | Co | Mn |
|----|-----------------|--------------|-----|------|------|------|------|-----|-----|
| 1 | 0 - 5 | 0.0 | 1.5 | 0.8 | 0.1 | 0.2 | 0.1 | 0.3 | 7.1 |
| 2 | 5 - 15 | 0.0 | 1 | 1.5 | 0.1 | 0.2 | 0.1 | 0.3 | 5.1 |
| 3 | 0 - 5 | 5.0 | 0.9 | 3.9 | 0.07 | 0.2 | 0.09 | 0.1 | 4.5 |
| 4 | 5 - 15 | 5.0 | 1.3 | 5.6 | 0.09 | 0.2 | 0.1 | 0.3 | 5.5 |
| 5 | 0 - 5 | 10.0 | 0.6 | 9.5 | 0.08 | 0.1 | 0.09 | 0.1 | 3.4 |
| 6 | 5 - 15 | 10.0 | 0.9 | 8.7 | 0.08 | 0.1 | 0.08 | 0.1 | 0.7 |
| 7 | 0 - 5 | 20.0 | 0.8 | 9.6 | 0.09 | 0.1 | 0.07 | 0.1 | 2.3 |
| 8 | 5 - 15 | 20.0 | 0.9 | 10 | 0.07 | 0.09 | 0.08 | 0.1 | 1.3 |
| 9 | 0 - 5 | 50.0 | 0.4 | 11.1 | 0.06 | 0.1 | 0.1 | 0.2 | 2.9 |
| 10 | 5 - 15 | 50.0 | 0.6 | 10.3 | 0.06 | 0.1 | 0.08 | 0.1 | 0.8 |
| 11 | 0 - 5 | 100.0 | 0.6 | 15.2 | 0.06 | 0.09 | 0.1 | 0.2 | 2.6 |
| 12 | 5 - 15 | 100.0 | 0.9 | 12.7 | 0.09 | 0.09 | 0.08 | 0.1 | 0.5 |

Table 7. Average elemental contents ($\mu\text{g g}^{-1}$ dm) of tomato roots and leaves grown adjacent to the national highway of Lamias-Athens Greece

| Distance (m) | Roots | | | | | | |
|--------------|-------|--------|-----|-------|------|------|-------|
| | Pb | Al | Cd | Zn | Cu | Co | Mn |
| 0 | 0.6 | 1741.9 | 0.8 | 31 | 14 | 13.5 | 159.4 |
| 5 | 0.8 | 545.2 | 0.8 | 30.7 | 11.9 | 5.3 | 77.4 |
| 10 | 1.1 | 1077.6 | 0.4 | 21 | 10.8 | 11.4 | 149.6 |
| 20 | 0.1 | 732.2 | 0.6 | 22.5 | 10.4 | 7.5 | 101 |
| 50 | 0.6 | 595.2 | 0.9 | 28.5 | 18.7 | 11.4 | 105.6 |
| 100 | 0.9 | 403.3 | 0.2 | 30.7 | 10.5 | 4.4 | 56.6 |
| Leaves | | | | | | | |
| | Pb | Al | Cd | Zn | Cu | Co | Mn |
| 0 | 0.9 | 143.2 | 1 | 103.5 | 8 | 2 | 117.5 |
| 5 | 0.4 | 90 | 0.3 | 40.2 | 12.8 | 1.2 | 72.8 |
| 10 | 0.6 | 61 | 0.6 | 22.5 | 12.3 | 1.3 | 87.4 |
| 20 | 0.6 | 92 | 0.3 | 62.9 | 10 | 1.8 | 85 |
| 50 | 0.9 | 87.1 | 0.5 | 42.8 | 11.6 | 2.2 | 75.6 |
| 100 | 0.7 | 74.3 | 0.5 | 50.1 | 11.3 | 2.0 | 74.8 |

Table 8. Regression models for chemical elements as functions of distance

| $D \rightarrow E_i (S)$ | $D \rightarrow E_i (R)$ | $D \rightarrow E_i (L)$ |
|---|---|--|
| $Pb_S = 1.1715 - 0.023524 * D + 0.00019369 * D^2$ | $Pb_R = 0.77274 - 0.013216 * D + 0.00014653 * D^2$ | $Pb_L = 0.60961 + 0.0052837 * D - 0.000041088 * D^2$ |
| $Al_S = 3.7561 + 0.25866 * D - 0.0016075 * D^2$ | $Al_R = 1241.4 - 23.820 * D + 0.15769 * D^2$ | $Al_L = 107.55 - 1.0683 * D + 0.0076740 * D^2$ |
| $Cd_S = 0.093417 - 0.0011230 * D + 0.0000093858 * D^2$ | $Cd_R = 0.63131 + 0.0078292 * D - 0.00011795 * D^2$ | $Cd_L = 0.67195 - 0.011872 * D + 0.00010477 * D^2$ |
| $Zn_S = 0.18184 - 0.0035369 * D + 0.000026742 * D^2$ | $Zn_R = 28.139 - 0.15851 * D + 0.0019111 * D^2$ | $Zn_L = 67.119 - 1.1173 * D + 0.0096719 * D^2$ |
| $Cu_S = 0.093953 - 0.00046843 * D + 0.0000044483 * D^2$ | $Cu_R = 10.960 + 0.18964 * D - 0.0018845 * D^2$ | $Cu_L = 10.375 + 0.048262 * D - 0.00039770 * D^2$ |
| $Co_S = 0.22126 - 0.0045690 * D + 0.000039747 * D^2$ | $Co_R = 9.5502 + 0.043880 * D - 0.00091509 * D^2$ | $Co_L = 1.4742 + 0.016801 * D - 0.00011160 * D^2$ |
| $Mn_S = 5.0046 - 0.13620 * D + 0.0010379 * D^2$ | $Mn_R = 129.37 - 0.62725 * D - 0.00081138 * D^2$ | $Mn_L = 98.018 - 0.84186 * D + 0.0061985 * D^2$ |

Analysis based on Qualitative Descriptions

With 5 differences for distance values, we can qualitatively describe the above factors by merging three adjacent cells or by viewing them as a qualitative evolution operator, depending on the task. The two-order descriptions are summarized in Table 9.

The most characteristic behavior features based on complete similarity (identity) are shown in Table 10. Also, identical case for the whole chain $S \rightarrow R \rightarrow L$ is **Zn**. The complete distinction cases are **Pb** for $R \rightarrow L$, **Al** and **Cu** for $S \rightarrow R$.

Table 9. Qualitative representation of relationships (24) with two-order descriptions

| | $D \rightarrow E_i (S)$ | $D \rightarrow E_i (R)$ | $D \rightarrow E_i (L)$ |
|-----------|-------------------------|-------------------------|-------------------------|
| Pb | 4+, +2+ | 4+, +=+ | 4+, -2+ |
| Al | 5+, -2+ | 5-, +2+ | 4+, +2+ |
| Cd | 4+, +2+ | +2+, -2- | 4+, +2+ |
| Zn | 4+, +2+ | 4+, +2+ | 4+, +2+ |
| Cu | 4+, +2+ | 4+, -2+ | +2+, -2+ |
| Co | 4+, +2+ | 3+2-, -2- | 5+, -2+ |
| Mn | 4+, +2+ | 5-, 4- | 4+, +2+ |

Table 10. The characteristic first-order descriptions

| First order description | Soil | Roots | Leaves |
|-------------------------|-------------------------------|---------------|-----------------------|
| (4→) | Pb, Cd, Zn, Cu, Co, Mn | Pb, Zn | Al, Cd, Zn, Mn |
| (5→) | | Al, Mn | |
| (5+) | Al | | Co |

It is also reasonable to analyze the behavior features according to partial similarity/distinction. Here, we can use the following two ways to analyze features with different degrees of similarity.

1) Sign-by-sign matching between two descriptions

When matching (\Leftrightarrow) two descriptions sign-by-sign, distinct signs of the same location are replaced with "*", e.g.: (4→) \Leftrightarrow (5+) = (4*+). The measure of similarity/distinction could be the ratio of number of identical signs to total number of signs. For that example, the similarity degree is $m=4/5$, and distinction degree is $d = 1/5$.

In particular, this way makes it possible to compare descriptions with a template representing features of interest, and to rank the descriptions according to the m or d values. The example of such ranking for the template (5+) is shown in Table 11.

Table 11. Example of descriptions ranking for the template (5+)

| | $q^1 (D \rightarrow E_i (R))$ | (5+) $\Leftrightarrow q^1$ | m |
|-----------|-------------------------------|----------------------------|-----|
| Cu | 4+→ | 4*+ | 4/5 |
| Cd | +→2+→ | +*2*+ | 3/5 |
| Co | 3+2→ | 3+2*+ | 3/5 |
| Pb | 4+→ | 4*+ | 1/5 |
| Zn | 4+→ | 4*+ | 1/5 |
| Al | 5→ | 5* | 0 |
| Mn | 5→ | 5* | 0 |

It is also possible to compare elements with the same rank, e.g.:

$$(\mathbf{Cd}, \mathbf{Co}, 3/5): (+\rightarrow 2+\rightarrow) \Leftrightarrow (3+2\rightarrow) = (+*+*-\rightarrow) \quad (25)$$

The (+*+*-) represents identical and distinct parts for two elements over the distance range. There is common increase of concentrations at the short and middle distances, and common decrease at the long distance.

For the pair (**Pb**, **Zn**, 1/5), the elements are completely similar.

2) Feature matching (max-min, location of peaks over D range, etc.)

The results of matching according to number of peaks are shown in Table 12. The peaks are determined simply according to change of adjacent signs: $(s_i^1, s_{i+1}^1) = (+ - / - +)$.

It is also possible to analyze elements with the same number of peaks, e.g. maxima for **Cu** and **Co**:

$$\mathbf{Cu} (++++\rightarrow) \Leftrightarrow \mathbf{Co} (++++-\rightarrow) = (++++*-\rightarrow) \quad (26)$$

Here, "*" shows a "floating" maximum at the fourth position. The sign-by-sign matching ($m(\mathbf{Cu}) = 4/5$ and $m(\mathbf{Co}) = 3/5$) together with the feature matching show that there is a close similarity between those elements for roots.

Assumptions based on qualitative analysis

The degree of similarity or distinction between concentrations of individual chemical element involved in direct factors, i.e. $S \rightarrow R$ and $R \rightarrow L$, can be used to make assumptions about degree of influence of other unaccounted factors (both elements and objects). This will make it possible to provide initial assessment of what kind of influence is stronger: the element itself or other factors. Those assumptions can be a basis to direct future experiments towards resolution of uncertainties about behavior of objects / elements.

Table 12. Example of descriptions ranking according to number of peaks

| | $D \rightarrow E_i(R)$ | Peaks | | |
|-----------|------------------------|-------|-----|-------|
| | | Max | Min | Total |
| Cd | +---+ | 2 | 1 | 3 |
| Pb | -----+ | 0 | 1 | 1 |
| Zn | -----+ | 0 | 1 | 1 |
| Cu | ++++- | 1 | 0 | 1 |
| Co | +++-- | 1 | 0 | 1 |
| Al | ----- | 0 | 0 | 0 |
| Mn | ----- | 0 | 0 | 0 |

For the relationship $S \rightarrow R$, complete similarity (according to the first-order differences) can be observed for **Pb**, **Zn**, quasi complete similarity (with one sign distinction) for **Mn**, whereas complete distinction for **Al**, **Cu**, and quasi complete distinction for **Co**.

For the relationship $R \rightarrow L$, complete similarity can be observed for **Zn**, quasi complete similarity for **Al**, **Mn**, and complete distinction for **Pb**.

For the whole chain $S \rightarrow R \rightarrow L$, complete similarity can be observed for **Zn**, quasi complete similarity for **Mn**, whereas complete or quasi complete distinction is not present.

Proceed from the above we may assume that concentrations of **Pb**, **Zn**, **Mn** in soil are substantial factors for the concentrations of the same elements in roots, while concentrations of **Al**, **Cu**, **Co** are not substantial factors. Similar assumptions can be made for $R \rightarrow L$ and for the whole chain.

5. CONCLUSIONS

The paper describes an approach to qualitative analysis of experimental results. Both examples are given for the initial stage of experimental study with very short data samples where it is impossible to carry out a comprehensive analysis involving large number of variables.

In order to overcome the problem of great number of variables with short data samples, two main points were taken into account: a) system decomposition into subsystems, enabling us to model and analyze individual subsystems with sufficiently small number of variables; b) integration and generalization of modeling results for different subsystems to the whole system.

Decomposition of system into subsystem allowed us to reduce number of variables, but we were able to satisfactorily estimate binary relationships only. The qualitative analysis technique gives the possibility to analyze behavior of individual subsystems as well as the system in whole.

The paper describes the application of Qualitative Description that includes signs combinations of differences obtained for a regression function. By definition of the difference itself, its sign reflects various states of mutual position of neighbor points. Therefore, the Qualitative Description can identify local functional features such as extrema, their forms, tendencies of change of features over time, etc. In case if analysis of absolute values is required, the Qualitative Description can be used together with qualitative variables like "high", "low". etc. (Kalavrouzotis *et al.*, 2006).

In particular, the qualitative analysis has shown that **P** concentration in Broccoli (*Brassica Oleracea var. Italica*) roots, falls for both *CW* and *WW*, but faster for *WW*, "faster" being derived from "-" for (*WW-CW*). Also, the **P** concentration in leaves falls for both *CW* and *WW* but more slowly for *WW*, "more slowly" being derived from "+" for (*WW-CW*). Similarly, with respect to the experimental data of tomato plants (*Lycopersicum esculentum*) cultivated near the highway, there is common increase of concentrations at the short and middle distances, and common decrease at the long distance. It is also assumed that the concentrations of **Pb**, **Zn**, **Mn** in soil are substantial factors determining the concentrations of those elements in roots of tomato plants, while the concentrations of **Al**, **Cu** and **Co** are not substantial factors. Such kind of assumptions can be used to design the next-stage experiments and obtain more comprehensive models of plant species behaviors.

At the early stages of experiments with very small samples, the GMDH-like methods can be applied for simple subsystems only. The system decomposition into big number of subsystems makes it difficult to analyze the system in whole. Therefore, application of the qualitative analysis tool is especially useful at the early stages.

The paper demonstrates the possibilities of the proposed qualitative analysis technique to identify main behavior features of individual objects/relationships of system/subsystem, classify behaviors according to different criteria of similarity/distinction, provide feature matching, etc. The approach to the representation of qualitative features and main operations are taken from the classifier systems (Booker *et al.*, 1989). Here, the paper does not consider the machine learning issues, focusing on the possibilities to provide the researcher with a qualitative analysis tool capable of suitably manipulating the experimental results. In practice, the researcher, who deals with the great number of subsystem plots, faces the problem to understand "the picture in whole", by trying to match features of different plots, generalize the results, etc., out of the computer-aided tools. With new stages of experiments and data accumulation, interpretation of the results becomes more and more difficult. The qualitative analysis tool gives the researcher a total though "shallow" picture. History of the QD analysis results can be written in a computer database, which provides a qualitative view of all-stage experimental study results.

REFERENCES

- Booker L.B., Goldberg D.E. and Holland J.H. (1989). Classifier systems and genetic algorithms, *Artificial Intelligence*, **40**, 235-292.
- Bratko I. and Suc D. (2003). Qualitative Data Mining and Its Applications, *Journal of Computing and Information Technology*, **11 (3)**, 145-150.
- De Kleer J. and Brown J.S. (1984). A qualitative physics based on confluences, *Artificial Intelligence*, **24**, 7-83.
- Drakatos P.A., Vissikirsky V.A., Kalavrouziotis I.K., Skuras, D.G. and Drakatu S.P. (2001). Usage of Fuzzy Modeling for Cultivation of Forests Irrigated by Wastewater Treatment Plants, *International Journal of Environmental Studies*, **58**, 813-833.
- Farlow S.J. (Ed.) (1984). *Self-organizing methods in modelling: GMDH type algorithms*, New-York, Bazel: Marcel Decker, 350 p.
- Kalavrouziotis I.K., Stepashko V.S., Vissikirsky V.A. and Drakatos P.A. (2002). Group Method of Data Handling (GMDH) application for modeling of mechanical properties of trees irrigated with wastewater, *International Journal of Environment and Pollution*, **18 (6)**, 589-601.
- Kalavrouziotis I.K., Stepashko V.S., Vissikirsky V.A. and Drakatos P.A. (2004). Group Method of Data Handling (GMDH) in modelling of growth dynamics of tress irrigated with wastewater, *International Journal of Environment and Pollution*, **21 (4)**, 350-364.
- Kalavrouziotis I.K., Pantazis V.N., Vissikirsky V.A. and Stepashko V.S. (2006). Irrigation of Forest Species with Wastewater and Sludge: Experiments, Modeling, and Qualitative Assessment, *Instrumentation Science and Technology*, **34**, 727-742.
- Kalavrouziotis I.K., Jones P.W., Carter J. and Varnavas S.P. (2007a). Uptake of trace metals of *Lycopersicum Esculendum* (Tomato) at a site adjacent to a major road, Lamias, Greece, *Fresenius Environmental Bulletin*, **16 (2)**, 133-139.
- Kalavrouziotis I.K., Koukoulakis P.H., Robolas P., Papadopoulos A.H. and Pantazis V. (2007b). Interrelationship of heavy metals and properties of a soil cultivated with *Brassica oleracea* var. *Italica* (Broccoli), under the effect of Treated Municipal Wastewater, *International Journal Air, Soil and Water Pollution*, **190**, 309-321.
- Kalavrouziotis I.K., Robolas P., Koukoulakis P.H. and Papadopoulos A.H. (2008). Effects of municipal reclaimed wastewater on the macro and microelements status of soil and of *Brassica oleracea* var. *Italica*, and *Brassica oleracea* var. *gemmifera*, *International Journal Agricultural Water Management*, **95**, 419-426.
- Kuipers B. (1984). *Qualitative reasoning: modeling and simulation with incomplete knowledge*. The MIT Press, Cambridge, Massachusetts, 452 pp.
- Kuipers B. (1986). Qualitative simulation, *Artificial Intelligence*, **29**, 289-338.
- Srivastava P.C., Gupta U.C. (1996). Trace Elements in Crop Production. Science Publishers Inc.10 Water Street, 310, Lebanon, NH 03766 USA, 209 p.
- Vissikirsky V.A. (2001). An Approach to the Organization of Qualitative Knowledge Proceed from Computer Simulation Results. Proceedings of ASME International - Greek Section, First National Conference on Recent Advances in Mechanical Engineering, Patras, Greece, 17-20 September.
- Vissikirsky V.A., Stepashko V.S., Kalavrouziotis I.K. and Drakatos P.A. (2005). Growth Dynamics of Trees Irrigated with Wastewater: GMDH Modeling, Assessment, and Control Issues, *Instrumentation Science and Technology*, **33**, 237-257.

- Vissikirsky V.A., Stepashko V.S. and Kalavrouziotis I.K. (2007). Application of GMDH to the Environmental Modeling with Short Samples, *Proceedings of International Workshop on Inductive Modeling (IWIM 2007), Prague, 22-26 September*, 18-25.
- Vissikirsky V.A., Stepashko V.S., Kalavrouziotis I.K. and Varnavas S.P. (2008a). The Road Pollution Impact on Zea mays: Inductive Modeling and Qualitative Assessment, *Water, Air and Soil Pollution*, **195**, 301-310.
- Vissikirsky V.A., Stepashko V.S. and Kalavrouziotis I.K. (2008b). GMDH-based decision-making support for ecological processes. Proceedings of the 2-nd International Conference on Inductive Modeling (ICIM 2008), Kyiv, 15-19 September, 310-313.