Application of GMDH to the Environmental Modeling with Short Samples

Vladimir A. Vissikirsky[1], Volodymyr S. Stepashko[1], and Ioannis K. Kalavrouziotis[2]

[1]*Department for Information Technologies of Inductive Modeling, the International UNESCO Center of Information Technologies and Systems of the National Academy of Sciences of Ukraine, Ukraine*
[2]*Department of Environmental and Natural Resources Management , University of Ioannina, Greece*

vis@mathmod.kiev.ua, astrid@irtc.org.ua, ikalavru@cc.uoi.gr

Abstract. *The paper considers the issues concerning the environmental modeling and analysis of agricultural and forest plant species grown under different controlled and uncontrolled conditions, with short data samples available. The key approach to the analysis of such kind of complex systems consists in the decomposition of a modeled system into subsystems with consistently small number of samples. The paper shows the examples of how to analyze individual subsystems and generalize modeling results for different subsystems to the whole system, by utilizing the GMDH advantages and qualitative analysis techniques. Here, the GMDH has been applied to estimate and analyze features of the tree growth dynamics, heavy metals impact on species cultivated near the motor road, etc. The GMDH generates the sets of models with different structures and accuracies thus enabling the identification of the most relevant structures, the analysis of their similarity and the assumptions about input of different factors. In particular, it is shown how a pair of linear models of identical structure describing different subsystems can be compiled into a model that describes a higher-level subsystem. Qualitative analysis techniques give the possibility to analyze behavior of individual subsystems as well as the system in whole.*

Keywords

GMDH, qualitative assessment, environmental modeling,
irrigation, wastewater, road pollution

1 Introduction

The paper considers the issues associated with environmental modeling and analysis under the conditions of short samples available. The results presented in the paper are based on the studies carried out to estimate characteristics of different plant species irrigated with processed wastewater, as well as heavy metals impact on species cultivated near the motor road. For these tasks, GMDH methods have been applied to estimate various aspects of plant species behavior: mechanical properties of trees (axial compression and bending) depending on concentrations of chemical elements contained in the wastewater [1], growth dynamics of trees irrigated with wastewater (development of tree height and mortality over time) [2, 3], etc. In addition, qualitative analysis techniques were applied to assess plant species behavior under different conditions: irrigation of forest species with wastewater and sludge [4], heavy metals impact on Zea Mays species cultivated near the major road in the area Araxos, Greece [5].

Short data samples available for the analysis stems from the nature of problems considered, in particular:

1) Growth of plant species is a slow process. Therefore, experiments designed to identify species behavior are conducted by series, during long-term period. Design of the next-stage experiments

implies making the decisions about whether to continue the experiments and what kind of experiments shall be completed to decrease the uncertainty degree of knowledge available.

2) There are a great number of input variables (e.g., concentrations of chemical elements) with small number of measurements.

Here, the key approach to resolve the problem consists in the decomposition of a modeled system into subsystems, to provide the modeling and analysis of subsystem(s) that have subset(s) of input variables consistent with volume of sample available. This approach is natural for the analysis of complex systems, which makes it possible to analyze in details different aspects of system behavior. At the same time, there is a need to coordinate and integrate the analysis of estimated sub models. Due to short data sample available at initial stage of system development, the initial analysis results are virtually **assumptions** that serve as a basis to design the next-stage experiments directed to decrease the uncertainty level or confirm preliminary conclusions.

We consider here two approaches allowing us to analyze both separate subsystems and the system in whole: structural and qualitative analyses based on the GMDH methods.

2 Structural Analysis Based On The Set Of Generated GMDH Models

Let us consider the decomposition of a system into subsystems, the subsystem model design, the merging of two subsystems (sub models) into a higher-order subsystem, as well as control and optimization issues.

The approach is based on the structural analysis of the set of GMDH generated models. The GMDH generates the set of polynomial dependencies with different degree of complexity and accuracy. Therefore, we can consider the set of generated models in whole and select suitable models according to the structure, complexity, and accuracy requirements. Although the most accurate model is preferred, analysis of the accuracies distribution of generated models and their structures gives the possibility to obtain additional information that allows us to make assumptions about input of different variables.

Let us consider the structural analysis on the example of growth dynamics models designed for the forest tree Pinus Brutia irrigated with different wastewater treatment conditions. The experiments were carried out to identify the height and mortality rates of the tree [2, 3].

The land plot was divided into four areas, according to different irrigation (treatment) cases applied to the seedlings during their growth:

Treatment case 1: Irrigation with waste water without application of sludge (*WW*);

Treatment case 2: Irrigation with control water without application of sludge (*CW*);

Treatment case 3: Irrigation with waste water and application of sludge (*SWW*);

Treatment case 4: Irrigation with control water and application of sludge (*SCW*).

The tree growth characteristics considered are the “tree height” and “tree mortality” for the above four treatment cases.

The process of tree species growth is considered as a time-dependent series of changes in the tree characteristics. In general, the tree species growth at a period (*t+1*) is a function of all *t* preceding growth periods. The growth dynamics model can be viewed as a model with distributed lagged variables. Here, for the assessment purposes, we will analyze the growth effects with one lag, i.e., the effects taking place during a current half-year period and one lagged half-year period, in order to take into account possible differences in climatic conditions during the half-year periods “spring-summer” and “fall-winter”. Therefore, the model will be the first-order lagged model with two output variables $\vec{Y}(t) = [y^{(h)}(t), y^{(m)}(t)]$ for each treatment case, where $y^{(h)}(t)$ is the tree height, and $y^{(m)}(t)$ is the tree mortality. Meteorological factors $\vec{X}(t) = [x_1(t), x_2(t), x_3(t)]$ are: temperature ($x_1(t)$), relative humidity ($x_2(t)$), and cloudiness ($x_3(t)$).

Water content as a factor is not explicitly included into the growth dynamics model, because it is approximately constant over the survey period for the control and waste water. Therefore, we will design the models for each treatment case and compare the individual models.

The linear models of tree species growth dynamics can be represented as dynamic input-output models with first-order lagged variables:

$$\begin{aligned} & y^{(k,j)}(t+1) = a_0 + a_{y(t)}y^{(k,j)}(t) + a_{1,t+1}x_1(t+1) + a_{1,t}x_1(t) + \\ & + a_{2,t+1}x_2(t+1) + a_{2,t}x_2(t) + a_{3,t+1}x_3(t+1) + a_{3,t}x_3(t), \\ & t = 0,..,7, k = \ [h \text{ - height, } m \text{- mortality}], j = \ [WW, CW, \ SWW, \ SCW] \end{aligned} \quad (1)$$

For the purpose to analyze the overall range of linear models, we will use the combinatorial GMDH algorithm COMBI.

Structural analysis consists of the following stages: 1) evaluation of model accuracies; 2) analysis of model structures; 3) merging of two models (subsystems) into general model for a treatment type (see below); 4) assessment of optimal irrigation conditions for a treatment type.

2.1 Evaluation of model accuracies

In order to make decision if linear models can be used in the assessment of tree growth characteristics, we need first to check their accuracies. Then, the subset of selected linear models with the best accuracies will be used in further analysis.

On the basis of preliminary analysis, proceed from the application field and experience, and by comparing accuracies of linear and non-linear models (COMBI and MIA) we can specify an accuracy range to select a subset of generated models that will further be analyzed.

Preliminary selection of a subset of models and its analysis can be demonstrated on the example of models estimating the treatment case “waste water”. By utilizing COMBI with 8 terms in linear models, maximum number of generated models with different structures is 2^8-1=255. On the basis of experience, we will accept here the normalized range of deviation (*RD*) as a measure of accuracy. The *RD* is a percentage ratio “maximum deviation between measured and estimated values / difference between maximum and minimum measured values”.

The best and the worst models of the first 150 generated models for the “tree height” (which are potentially most representative with no degraded structure) have *RD* equal to 0.6076% and 3.9894% respectively. Proceed from that, one can accept $RD \le 2.5\%$ covering over 80% of the considered models. Similar consideration can be done for the tree mortality, waste water treatment case. These models are less accurate than those for the tree height. For the assessment purposes, let us accept $RD \le 30\%$ covering 75% of the considered models. Thus, the *RD* thresholds determine the subsets of models that will be used in further analysis. The threshold values at first step are approximate and subjective, and they can be refined at next stages within the iterative process “experiments-modeling-analysis”.

2.2 Analysis of model structures

Further analysis of model structures shall be done within the accuracy range according to the criteria of accuracy distribution and model structures.

Assessment of structures of model set. When considering the features of a single model, the existence of a set of models with acceptable accuracy and all positive coefficients indicates stable increase of an output value. Thus, there are 29% of models with positive coefficients for the “tree mortality”, treatment case “sludge and waste water”, for which $RD \le 9\%$ (without absolute term). Distribution of accuracies for the models with positive coefficients is better than that for the models having coefficients with unlike signs. No models exist with negative coefficients only. Therefore, one can assume that dominated is the case with general positive input of factors into the result.

Pairwise matching of model structures. Pairwise matching of structures of the models describing two different treatment cases is intended to assess the level of similarity / difference between model structures, hence, determine how two treatment cases are similar or different. The existence of identical or similar models for different treatment cases allows us to merge them into one generalized model describing a higher-level subsystem. Identical structures imply the existence of the same set of input variables and the same combination of signs of coefficients. In particular, the existence of all positive coefficients (without absolute term) such that all coefficients of identical terms of the models meet the same relation (> or <), leads to the assumption about stable difference between output values of models in the pair. Additional test for non-existence of a pair with similar structure but with inverse relation between coefficients of identical terms would confirm such assumption.

As an example, for the "tree height", treatment cases *WW* and *CW*, we can choose two pairs of models *1+2* and *3+4*, each model having all positive coefficients. The relations (>) are met between coefficients of the models in each pair. Here, as a measure of accuracy for a pair (RD_{total}), we will take the sum of *RD* values for each model in the pair. The pairs are shown below:

Treatment case WW, Model 1 (*RD*=3.9894%):

$$y^{(h,ww)}(t+1) = 1.0317\, y^{(h,ww)}(t) + 14.6450\, x_3(t) \tag{2}$$

Treatment case CW, Model 2 (*RD*= 3.5480%):

$$y^{(h,cw)}(t+1) = 1.0303\, y^{(h,cw)}(t) + 13.8030\, x_3(t) \tag{3}$$

with RD_{total} = 7.5374%, and

Treatment case WW, Model 3 (*RD*=1.6592%):

$$y^{(h,ww)}(t+1) = 0.9945\, y^{(h,ww)}(t) + 7.8314\, x_3(t+1) + 11.3570\, x_3(t) \tag{4}$$

Treatment case CW, Model 4 (*RD*= 1.0840%):

$$y^{(h,cw)}(t+1) = 0.9944\; y^{(h,cw)}(t) + 7.0173\, x_3(t+1) + 10.9390\, x_3(t) \tag{5}$$

with RD_{total} = 2.7432%

At the same time, there are no pairs with identical structure and inverse relations between coefficients (i.e. <). This shows that under the same initial conditions the height increase when irrigating with waste water is greater than the increase when irrigating with control water, and this difference is stable enough.

In general, the structure of a linear model *S* can be represented as an ordered set of symbols {1,0} indicating the existence or non-existence (1 or 0) of relevant term in the model, together with the accuracy. The structure of linear model (4) can be represented in the form: S_{13} = (0,1,0,0,0,0,1,1/1.6592).

2.3 Merging of two models into general model for a treatment type

One of the purposes of modeling and assessment is to generalize the models obtained for four treatment cases into the following two treatment types:

Treatment type 1 (*W*): irrigation with water without application of sludge (treatment cases *WW* and *CW*);

Treatment type 2 (*SW*): irrigation with water and with application of sludge (treatment cases *SWW* and *SCW*).

Each treatment case is associated with constant values of concentrations of chemical elements in the control or waste water. General model for a treatment type shall describe the growth dynamics at varying chemical content of irrigation water. In order to build a general model for a treatment type, *WW, CW -> W or SWW, SCW -> SW*, we shall associate the models with some indices of water content. This would enable us the explicit assessment of the effect of water quality on tree species

growth characteristics, as well as the solving of optimum value of water content maximizing the tree species growth by a given criterion.

Water index. The indices of irrigation water content, pH and concentrations of chemical elements in the control and waste water, are approximately constant for all periods. Since we have the averaged constant values of water content, let us define the overall water index Q characterizing water content in general, in order to further express the coefficients of general model as variables of the overall water index. For the assessment purposes we will accept an overall water index without "weights" assigned to individual chemical elements, and taking into account that all concentrations of elements in the control and waste water are less than the recommended maximum concentrations:

$$Q = \frac{1}{n} \sum_{i=1,n} \frac{(m_i - c_i)}{m_i} \tag{6}$$

Here, m_i is the recommended maximum concentration of *i-th* chemical element, c_i is the measured concentration of *i-th* chemical element. The range of such index is [0,1]. The limit 0 means that all measured concentrations are equal to the recommended maximum concentrations, and limit 1 means that all concentrations are equal to 0. The overall water index for control water is $Q^{(cw)} = 0.8330$, and for waste water is $Q^{(ww)} = 0.7942$, so in general waste water is more saturated.

Example of model merging. ***Treatment type 1: WW, CW -> W, general model for "tree height"***

Let us select the best pair from the subset of models with identical structure and positive coefficients, without absolute term. The best pair of models is *3+4* (see (4) and (5)). The general model is the following:

$$y^{(h,w)}(t+1) = a^{(h,w)}_{y(t)}(Q)\ y^{(h,w)}(t) + a^{(h,w)}_{3,t+1}(Q)\, x_3(t+1) + a^{(h,w)}_{3,t}(Q)\, x_3(t) \tag{7}$$

Since we have two values of Q for treatment cases *WW* and *CW*, the coefficients a_s of models can be represented as a linear dependence of Q:

$$a_s = b_{1,s} + b_{2,s} \cdot Q \tag{8}$$

With a greater number of Q values one can form the dependence of higher order.

Linear dependence for the coefficients is obtained by pair of values (Q, a_s) which are known for each treatment case. Thus, for the coefficient $a^{(h,w)}_{3(,t+1}(Q)$, the linear dependence can be built by the following points: $(Q^{(ww)}, a^{(h,ww)}_{3,t+1}) = (0.7942, 7.8314)$ and $(Q^{(cw)}, a^{(h,cw)}_{3,t+1}) = (0.8330, 7.0173)$. Then the coefficients can be expressed as:

$$a^{(h,w)}_{y(t)} = 0.9965\text{-}0.0026 \cdot Q;\ a^{(h,w)}_{3,t+1} = 24.4953\text{-}20.9820 \cdot Q;\ a^{(h,w)}_{3,t} = 19.9131\text{-}10.7732 \cdot Q, \tag{9}$$

2.4 Assessment of optimal irrigation conditions for a treatment type

The purpose of water control on the basis of estimated models is the regulation of water content in a way to provide the best characteristics of tree species growth. In this case the overall water index $Q(t)$ on the range [0.7942, 0.8330] is considered as a control variable. By optimal treatment condition for each treatment type we will understand such $Q(t)$ that maximizes "tree height" while minimizes "tree mortality". Hence, the general merged models can be used to find optimal treatment conditions.

Preliminary analysis of structures of two general models for treatment type *W* shows that maximum for "tree height" and minimum for "tree mortality" are reached at the low bound of the overall water index range [0.7942, 0.8330], i.e. $Q_{opt} = 0.7942$ (waste water) for all periods.

In case of treatment type *SW*, the optimal *Q* is within the range [0.7942, 0.8330], which can be found by successive approximation. The optimal *Q* will show the proportion in which *CW* and *WW* shall be mixed at different irrigation periods to reach the maximum result.

3 Qualitative Analysis Based On The GMDH Models

The considered task is typical for initial experimental stage for which it is required to make initial analysis of factors impact. The task concerns the estimation of effects of heavy metal concentrations in soil, leaves and roots of Zea Mays species cultivated near the major road in the area Araxos, Greece. [5]

Here, the objective of the GMDH modeling is to obtain the regression models describing different relationships within the chain of interactions "soil - roots - leaves", as well as to provide further qualitative analysis of involved factors. Qualitative assessment provides initial generalization of regularities in change of heavy metals concentrations in soil, roots, and leaves, depending on distance from the road. It also allows us to determine the level of certainty / uncertainty of the models and to direct future experiments to obtain more detailed information.

The studied system involves three objects: soil (*S*), roots (*R*), and leaves (*L*), and has the following structure:

$$S\,(D) \rightarrow R\,(D) \rightarrow L\,(D) \tag{10}$$

where *D* is the distance from the road. Concentrations of chemical elements were measured in soil, roots, and leaves at distances of 0, 10, and 20 m from the road.

For this structure, a contiguous relationship, e.g. *S (D) -> R (D)*, is the direct factor for which it is assumed that *S (D)* is the main object effecting on *R (D)*; *S (D)* is the indirect factor for *L (D)*. In real situation there may exist other factors unaccounted in (10), for example, air environment, rainfall, etc.

Since at initial stage of the study there is lack of data enabling the modeling of complex interactions of chemical elements with account of direct and indirect factors, we will design and analyze the regression models for individual elements as functions of distance, as well as estimate the degree of possible influence of other elements and factors.

Let $C_j \in$ {*S* - Soil, *R* - Roots, *L* - Leaves} is the set of studied objects, $E_i \in$ {Pb, Al, Cd, Zn, Cu, Ni, Co, Fe, Mn} is the set of chemical elements studied in the objects C_j. In general, estimation models have the following form:

$$D \rightarrow E_i\,(C_j) \tag{11}$$

i.e., change in concentration of each element E_i in soil, roots, and leaves depending on distance, and the form:

$$(D, E_i\,(C_j)) \rightarrow E_i\,(C_k) \tag{12}$$

Qualitative analysis of the functions $(D, E_i\,(C_j)) \rightarrow E_i\,(C_k)$ allows us to assess the direct effects of individual chemical elements, i.e. relationships between concentrations of an element within direct factors *S -> R* and *R -> L*. The parameter of distance included in the function is in fact the assessment of degree of influence of other 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. The models are analyzed on the basis of inclusion/non inclusion of the parameter *D*.

Let us consider the estimation of different chemical elements within the chain "roots -> leaves":

1) $\boldsymbol{Pb}_{\mathbf{L}} = 0.578 - 0.176 * \boldsymbol{Pb}_{\mathbf{R}} + 0.009 * (\boldsymbol{Pb}_{\mathbf{R}})^2$

2) $\boldsymbol{Cu}_{\mathbf{L}} = 10.227 - 0.369 * \boldsymbol{D} + 0.01 * (\boldsymbol{D})^2$

3) $\boldsymbol{Zn}_{\mathbf{L}} = 22.1 - 0.385 * (\boldsymbol{D})^2 + 0.343 * \boldsymbol{Zn}_{\mathbf{R}}$

Case 1 ($\boldsymbol{Pb}_{\mathbf{L}}$). The only parameter $\boldsymbol{Pb}_{\mathbf{R}}$ is included in the model; *D* is not present. Therefore, it is assumed that the element ***Pb*** in roots is the main factor effecting on ***Pb*** in leaves. Other possible factors are not substantial.

Case 2 (Cu_L). The only parameter *D* is included in the model; Cu_R is not present. Therefore, it is assumed that the element ***Cu*** in roots is not the main factor effecting on ***Cu*** in leaves. There are other factors not included in the models, which effect on ***Cu*** in leaves.

Case 3 (Zn_L). Both parameters, *D* and Zn_R, are included in the model. Concentration of the element ***Zn*** in leaves depends on both concentration of ***Zn*** in roots and other factors, individually or as a result of their interactions. There is uncertainty about what factor(s) is predominant.

To obtain the generalized qualitative picture, we can process the results in the following manner. Let us partition the distance *D* into two crisp distance intervals: *D = {D1= [0-10], D2= [10-20]}*. For each distance interval, corresponding qualitative values *QV* are assigned, which describe the change of concentrations in soil, roots, and leaves:

$$QV=\{Z, LD, MD, HD, LI, MI, HI\} \quad (13)$$

where: *Z*- No Change; *LD* - Small Decrease, *MD* - Medium Decrease, *HD* - High Decrease, *LI* - Small Increase, *MI* - Medium Increase, *HI* - High Increase. Qualitative values describing the change of concentrations of elements at different distance intervals can be summarized in the **Tab.1**.

Tab.1. Qualitative values obtained for soil, roots, and leaves

Elements	**Soil, QV_S**		**Roots, QV_R**		**Leaves, QV_L**	
	D1	***D2***	***D1***	***D2***	***D1***	***D2***
Pb	*HD*	*LD*	*HD*	*LI*	*LD*	*HD*
Al	*HD*	*Z*	*HD*	*LI*	*HD*	*LD*
Cd	*HD*	*MI*	*MD*	*HI*	*Z*	*HD*
Zn	*HD*	*LI*	*HD*	*LI*	*HD*	*LD*
Cu	*HD*	*LD*	*HD*	*Z*	*HD*	*LD*
Ni	*HD*	*LD*	*HD*	*LD*	*LI*	*HD*
Co	*HD*	*LI*	*HD*	*LI*	*Z*	*Z*
Fe	*HD*	*Z*	*HD*	*LI*	*HD*	*LD*
Mn	*HD*	*LD*	*HD*	*LI*	*MD*	*MD*

This representation facilitates general view of tendencies of change, as well as computer-based processing of different cases and subsets of elements. Lets us summarize the number of occurrences over the distance range and generalize the result in terms of the Decrease/Increase tendencies:

Soil: {(*D1*: 9 *HD*), (*D2*: 4 *LD*, 2 *LI*, 2 *Z*, 1 *MI*} or {(*D1*: 9 Decrease), (*D2*: 4 Decrease, 3 Increase};

Roots: {(*D1*: 8 *HD*, 1 *MD*), (*D2*: 6 *LI*, 1 *HI*, 1 *Z*, 1 *LD*} or {(*D1*: 9 Decrease), (*D2*: 7 Increase, 1 Decrease};

Leaves: {(*D1*: 4 *HD*, 2 *Z*, 1 *MD*, 1 *LD*, 1 *LI*), (*D2*: 4 *LD*, 3 *HD*, 1 *MD*, 1 *Z*} or {(*D1*: 6 Decrease, 1 Increase), (*D2*: 8 Decrease}.

The results can be represented in the **Tab.2**.

Tab.2. General tendencies of change

Distance / **Object**	***D1***	***D2***
Soil	Decrease	Decrease/Increase
Roots	Decrease	Increase
Leaves	Decrease	Decrease

From the Table above, it can be clearly seen that the overall tendency is the decrease of concentrations, except for roots at the *D2*, and altering tendencies for soil at *D2*.

At the next stage with larger sample volume, one can estimate more complicated regressions.

4 Conclusion

The purpose of the paper is to demonstrate the advantages of GMDH approach to model and analyze complex environmental systems at the initial stage of their study when the sample volume is minimal. The set of generated GMDH models gives the possibility to analyze the system from different viewpoints, according to the model accuracies and structures. In the decomposition of a modeled system into subsystems, the related sets of GMDH models can be analyzed to identify the most relevant structures and make assumptions about input of different factors, assess the uncertainty level regarding each factor input, as well as to make decision on how to design the next-stage experiments to obtain more information about system behavior. This also makes it possible to compare the sets of GMDH models describing different subsystems by their structural similarity / difference, hence, merge the models into a model that describes a higher-level subsystem.

In addition, this provides the basis to develop simple reasoning procedures for qualitative assessment of system behavior at different stages of long-term experiments. Qualitative analysis gives the possibility, in terms convenient for the end user, to analyze behavior of individual subsystems as well as of the system in whole.

References

[1] Kalavrouziotis I. K., Stepashko V. S., Vissikirsky V. A., Drakatos P.A.: Group Method of Data Handling (GMDH) application for modelling of mechanical properties of trees irrigated with wastewater. *International Journal of Environment and Pollution,* p. 589-601, 18 (6), 2002.

[2] Kalavrouziotis I. K., Stepashko V. S., Vissikirsky V. A., Drakatos P.A.: Group Method of Data Handling (GMDH) in modelling of growth dynamics of tress irrigated with wastewater. *International Journal of Environment and Pollution*, p. 350-364, 21 (4), 2004.

[3] Vissikirsky V.A., Stepashko V.S., Kalavrouziotis I.K., Drakatos P.A.: Growth Dynamics of trees Irrigated with wastewater: GMDH Modeling, Assessment and Control Issues. *Instrumentation Science and Technology*, p. 229-249, 33 (2), 2005.

[4] Kalavrouziotis I. K., Pantazis V. N., Vissikirsky V. A., Stepashko, V. S.: Irrigation of Forest Species with Wastewater and Sludge: Experiments, Modeling, and Qualitative Assessment, *Instrumentation Science and Technology*, p. 727–742, 34, 2006.

[5] Kalavrouziotis I. K., Carter J., Varnavas S. P., Mehra A., Drakatos P.A.: Towards an understanding of the effect of road pollution an adjacent food crops: Zea mays as an example. *International Journal of Environment and Pollution*, 29, 2007.