De-noise data by using Multivariate Wavelets in the Path analysis with application

Lec. Dr. Nazeera Sedeek Kareem

Assis. Prof. Dr.Taha H.A. AL Z. Asst. Lec. Awaz shahab M

Univ. of Salahaddin/ College Administration &Economic

تخفيض الضوضائية باستخدام المويجات متعددة المتغيرات في تحليل المسار مع التطبيق

م. د. نظيرة صديق كريم

م. م. آواز شهاب محمد ا.م. د. طه حسين علي الزبيدي

كلية الادارة والاقتصاد / جامعة صلاح الدين

تاريخ استلام البحث 2017/4/10 تاريخ قبول النشر 2017/5/9

Abstract:

In this research dealing with pollution in multivariate data used in the analysis path through the purification by an algorithm proposed based on multivariate wavelets (Daubechies, bior and rbio) and soft thresholding and estimate of level in a way (Minimax) and then compare them with the results path analysis before dealing with pollution through the practical application of analyzing the causes of a problem water pollution in the Kurdistan region, concluded research on the proposed algorithm is efficient compared with the classical method depending on the MATLAB language and program (SPSS) with (Amos).

Keywords: De-noise, multivariate Wavelets, Path analysis, thresholding.

المستخلص:

تم في هذا البحث معالجة مشكلة تلوث البيانات المستخدمة في تحليل المسار عندما تكون ذات أبعاد متعددة من خلال تنقيتها بخوارزمية مقترحة تعتمد على الموجات متعددة المتغيرات (Daubechies, bior and rbio) وعلى قطع العتبة الناعمة وتقدير مستواها بطريقة (Minimax) ومن ثم مقارنتها مع نتائج تحليل المسار قبل معالجة مشكلة التلوث من خلال تطبيق عملي تناول مسببات مشكلة تلوث المياه في إقليم كردستان، وأستنتج البحث على كفاءة الخوارزمية المقترحة مقارنة مع الطريقة التقليدية وذلك من خلال استخدام لغة ماتلاب والبرنامج الإحصائي الجاهز (SPSS) مع (Amos).

1. Introduction:

Sewall Wright (1918-1934) developed a method of estimating causal path coefficients by decomposing the correlations among a set of variables. He articulated a set of rules for examining a path diagram that would allow this mathematical decomposition. [Alwin & Hauser, (1975)] The correlation of any two variables in a path diagram can be expressed as the sum of coefficients that connect the two variables. The connection between one variable and another variable, then, can often be made through more than one route.De noising becomes an indispensable step prior to analysis data. Multivariate wavelet shrinkage techniques are particularly well suited for such de noising tasks because they can yield a sparse representation of the data. There are several good reasons why multivariate wavelet shrinkage can be used for function estimation. The main reasons are that wavelet shrinkage estimators are: nearly minimax for a wide range of loss functions and for general function classes; simple, practical and fast; adaptable to spatial and frequency in homogeneities; readily extendable to high dimensions; applicable to various other problems such as density estimation and inverse problems. A review of these reasons and justification for them appears in Donoho and Johnstone (1995a). And which can be used to treat noise in the data.The practical side data applied for different environmental impacts which have influences on the water pollution [Simeonov and et. el., (2003)] which can be defined in many ways usually, it means one or more substances have built up in water to such an extent that they because problems could affect the health of all the humans, animals, and plants. Measurements like those are known as chemical parameter refer to any minerals, salts, metals, or anions dissolved in water.

2. Methodology

2.1. Path analysis

Path diagrams are useful enough as simple descriptive devices, but they can be much more than that. [Loehlin & Beaujean, (2017)] Starting from empirical data, one can solve for a numerical value of each curved and straight arrow in a diagram to indicate the relative strength of that correlation or causal influence. Numerical values, of course, imply scales on which they are measured. For most of this research we assume that all variables in the path diagram are expressed in standard score form, that is, with mean zero and standard deviation one. Covariances and correlations are thus identical. This simplifies matters of presentation, and is a useful way of proceeding in many practical situations. Later, we see how the procedures can be applied to data in original raw score units, and consider some of the situations in which this approach is preferred.

Because path analysis is an application of multiple linear regression, the same assumptions apply. In addition, it is more important than in MLR to have multivariate normal distribution of all the variables. This assumption is particularly important for the more general version of path analysis: structural equation modeling [Rex, (2004)].

1. Path analysis will require as many multiple linear regression analyses as the number of endogenous variables in the diagram.
2. Linearity. Check with partial regression scatter plots. Transform as necessary.
3. Multivariate normality. Check univariate distributions for normality with Kolmogorov-Smirnov or other test. Use χ^2 test of Mahalanobis distances for multivariate normality. Use transformations as necessary.

4. Outliers. Examine scatter plots. Use Mahalanobis distance. Eliminate outliers, with precautions.

Path coefficient: A standardized regression coefficient (beta), showing the direct effect of an independent variable on a dependent variable in the path model. In continuous variables the following Model is assumed:

$$X_k = \sum_{i=1}^{k-1} \beta_i X_i + \varepsilon_k \qquad (k = 2,3,...,K) \qquad \cdots \quad (1)$$

Where the variables X_i re standardized and ε_k is error term independent of explanatory variables X_i. In this case the direct effect of X_i on X_k is defined by:

$$e_d(X_i - X_k) = \beta_i \qquad \cdots \quad (2)$$

And determination of coefficient of X_i on X_k is defined by:

$$R^2(X_i - X_k) = \beta_i^2 \qquad \cdots \quad (3)$$

Where ρ_{ij} are the correlation coefficient between X_i and X_j. In this case the interpretation of the effects is easy, and the indirect effect of X_i on X_k is defined by:

$$e_{ind}(X_i - X_k) = \sum_{j=i+1}^{k-1} \beta_j \rho_{ij} \qquad \cdots \quad (4)$$

2.2. Multivariate Wavelets

General multivariate periodic wavelets are an efficient tool for the approximation of multidimensional functions, which feature dominant directions of the periodicity. One-dimensional shift invariant spaces and tensor-product wavelets are generalized to multivariate shift invariant spaces on non-tensor-product patterns. [Wang and Rose, (2006)] In particular, the algebraic properties of the auto-orphism group are investigated. Possible patterns are classified. By divisibility considerations, decompositions of shift invariant spaces are given. The results are applied to construct multivariate orthogonal Dirichlet kernels and the respective wavelets.

Over years, significant research interests have focused on using the wavelet de-noising techniques in the univariate case, but much less attentions on the de-noising of multivariate time series data. The basic procedure of multivariate wavelet de-noising technique is as follows: [Kaijian and et. el., (2012)].

1. It firstly projects the original data series into different scales using wavelet transform. The resulting coefficients would include approximation coefficients as well as horizontal, vertical and diagonal directions.
2. For approximation and direction coefficients at each direction, the threshold is chosen specifically at different scales for different directions and the wavelet coefficients are processed by either suppression or shrinkage.
3. Using the de-noised wavelet coefficients and the scale chosen, the processed wavelet coefficients are reconstructed into the unified de-noised data series using wavelet synthesis.

In the multivariate setting, given the multivariate variables X, the de-noising algorithm assumes that it consists of both deterministic data D and undesirable stochastic noises ω in a linearly fashion. Applying the multivariate wavelet analysis, this relationship is defined as in (5).

$$O \equiv \upsilon \mathrm{X} = \upsilon \mathrm{D} + \upsilon \omega \qquad \cdots \quad (5)$$

Where υ is an $N \times N$ orthonormal matrix, O is the N dimensional vector of wavelet transform coefficients $O_l : l = 0, \cdots, N-1.$

The soft thresholding focus on the signal smoothing [Donoho and Johnstone, (1995)]. It suppresses the wavelet coefficients below the set threshold value and subtracts the threshold value from the remaining wavelet coefficients. Compared with the hard thresholding, the data processing following soft threshold selection rules are smoother but lost the abrupt changes in the original data. It filters the signal as in (6):

$$O_l^{(st)} = sign(O_l)(|O_l| - \delta)_+ \qquad \cdots \quad (6)$$

Where

$$sign(O_l) = \begin{cases} +1 & if \quad O_l > 0 \\ 0 & if \quad O_l = 0 \\ -1 & if \quad O_l < 0 \end{cases} \qquad \cdots \quad (7)$$

And

$$(u)_+ = \begin{cases} u & if \quad x \geq 0 \\ 0 & if \quad x < 0 \end{cases} \qquad \cdots \quad (8)$$

Minimax threshold is one of the commonly used thresholds. The Minimax threshold is defined as threshold δ which minimizes the expression:

$$\inf_{\delta} \sup_{\theta} \left\{ \frac{R_{\delta}(\theta)}{n^{-1} + \min(\theta^{2},1)} \right\} \quad \cdots \quad (9)$$

Where $R_{\delta}(\theta) = E(\lambda_{\delta}(O) - \theta)^{2}$, $O \sim N(\theta,1)$

On the other hand it will be explained wavelets used in the search, as follows:

Daubechies wavelets have a support of minimum size for any given number N of vanishing moments. [Andreas, (2006)] Daubechies wavelet of class D-2N is a function $\Psi = {}_{N}\Psi \in L^{2}(R)$ defined by:

$$\psi(x) := \sqrt{2} \sum_{k=0}^{2N-1} (-1)^{k} h_{2N-1-k} \varphi(2x-k) \quad \cdots \quad (10)$$

Where $h_0, \cdots, h_{2N-1} \in R$ are the constant filter coefficients satisfying the conditions:

$$\sum_{k=0}^{N-1} h_{2k} = \frac{1}{\sqrt{2}} = \sum_{k=0}^{N-1} h_{2k+1}$$

As well as, for $l = 0,1,\cdots,N-1$

$$\sum_{k=2l}^{2N-1+2l} h_{k} h_{k-2l} = \begin{cases} 1 & if \quad l = 0 \\ 0 & if \quad l \neq 0 \end{cases}$$

And where $\varphi = {}_{N}\varphi : R \to R$ is the (Daubechies) scaling function (sometimes also “scale” or “father wavelet”), given by the recursion equation

$$\varphi(x) = \sqrt{2} \sum_{k=0}^{2N-1} h_{k} \varphi(2x-k) \quad \cdots \quad (11)$$

Bior-3.3 (Biorthogonal Wavelets) it is well known that bases that span a space do not have to be orthogonal. In order to gain greater flexibility in the construction of wavelet bases, the orthogonality condition is relaxed allowing semi-orthogonal, biorthogonal or non-orthogonal wavelet bases. Biorthogonal Wavelets are families of compactly supported symmetric wavelets. The symmetry of the filter coefficients is often desirable since it results in linear phase of the transfer function. In the biorthogonal case, rather than having one scaling and wavelet function, there are two scaling functions φ an d $\tilde{\varphi}$ that may generate different multiresolution analysis, and accordingly two different wavelet functions ψ and $\tilde{\psi}$ is used in the analysis and Ψ is used in the synthesis. And rbio wavelet is reverse biorthogonal has the Properties: symmetric, not orthogonal, biorthogonal.

2-3: Proposed Method:

The data will be contamination treatment (all the independent variables and the dependent variable together) before the path analysis through purification depending on three multivariate wavelets (Daubechies from 8-order, bior-3.3 and rbio-3.7) and get the multivariate (DWT) then involve the thresholding and estimation level of using Minimax threshold to obtain a modified multivariate (DWT), and taking its inverse we get on de-noise data, the following diagram shows the proposed method:

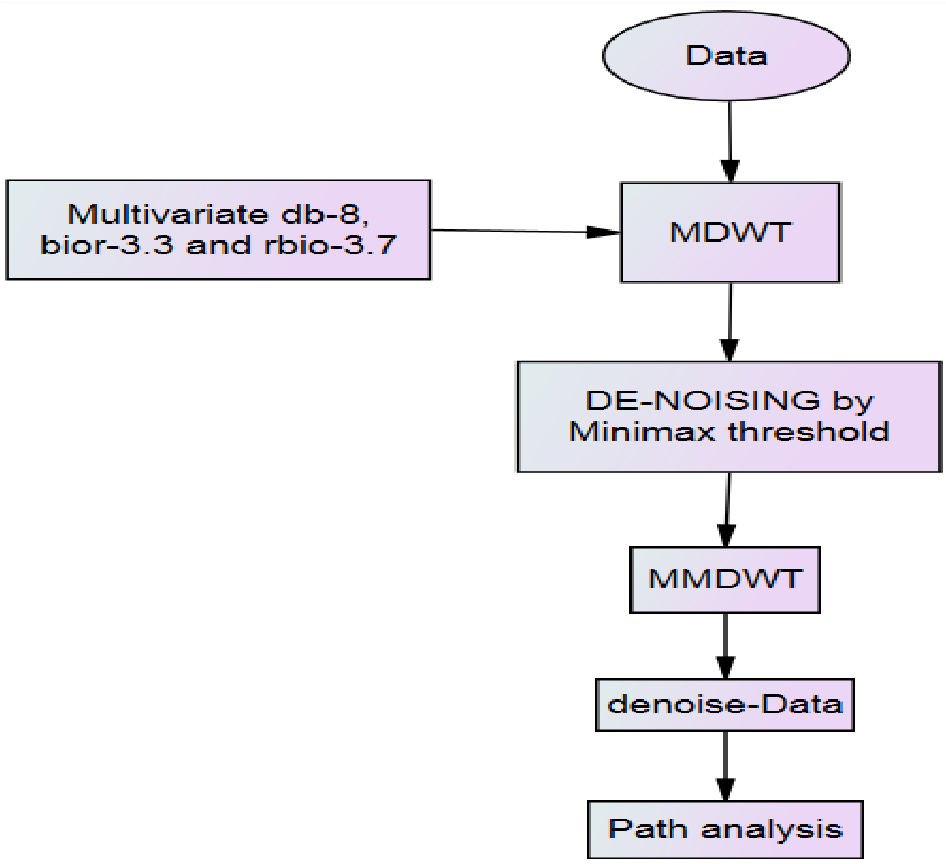

Figure (1)
Proposed method for path analysis

3: Practical side

The present study deals with the assessment of chemical characteristics of Dokan, Derbendikhan and Duhok three lakes in Kurdistan. chemical characteristics of these lakes have been studied and analyzed for a year, during January- April-July and October 2009. Various parameters including water pollution decomposition of TDS, as a dependent variable (Y), and independent variables including calcium (Ca), magnesium (Mg), sulfate (SO_4), bicarbonate (HCO3) and nitrite (NO3) have been analyzed**.** The research depends on collecting and analyzing 20 samples of water from selected areas.

With path analysis, multiple regression is used in conjunction with a causal theory, with the aim of describing the entire structure of linkages between dependent and independent variables posited from that theory, based on theoretical considerations of the Y (see Appendix table (A)), the researcher has constructed the path model presented in Figure (1) to represent the hypothesized structural relationships between the five variables Ca, HCO3, SO4, NO3 and Mg.

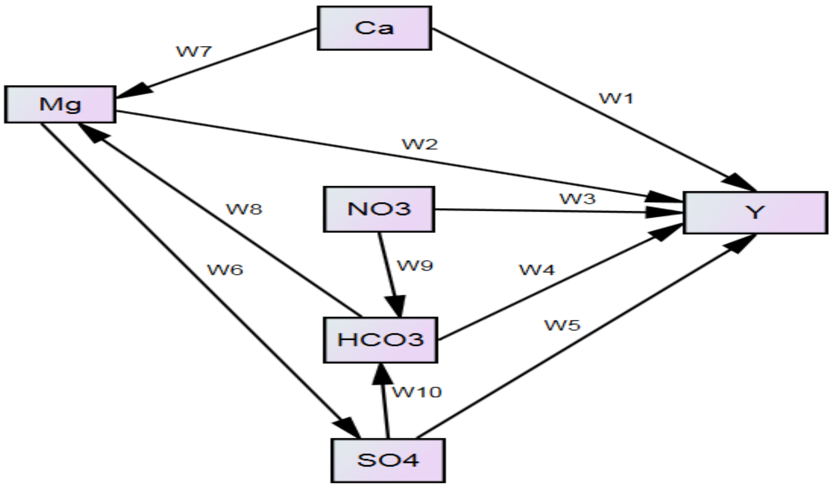

Figure (2)

the hypothesized structural relationships

The model specifies an “ordering” among the variables that reflects a hypothesized structure of cause-effect linkages. Multiple regression technique can be used to determine the magnitude of *direct* and *indirect* influences that each variable has on other variables that follow it in the presumed causal order (as indicated by the directional arrows). Each arrow in the model represents a presumed causal linkage or path of causal influence. Through regression techniques, the strength of each separate path can be estimated. This analysis actually involves four regression equations because Model (1) the response variable (Y) is a dependent variable on the five independent variables, Model (2) SO4 variable is a dependent variable for one variable of Mg, Model (3) the Mg variable is a dependent variable for two variables of Ca and HCO3, and Model (4) the HCO3 variable is a dependent variable for two variables of NO3 and SO4.

3-1: Classical Method

Check univariate distributions for normality with One-Sample Kolmogorov-Smirnov test and summarizes the most important results through the following table:

Table (1)
Kolmogorov-Smirnov Test (Original data)

	Y	Mg	SO4	HCO3
Kolmogorov-Smirnov Z	.861	1.008	.712	.809
Asymp. Sig. (2-tailed)	.448	.261	.691	.529

The results show that all p-values are greater than 0.05 which means that all the variables (dependent variables for four models) have a univariate normal distribution.

Using a χ^2 test was all values of Mahalanobis distances (for multivariate normality) less than tabulated χ^2 under the significant level of 1% and 4 degrees of freedom equal to (13.28) which means that the four variables (dependent variables for four models) have a multivariate normal distribution.

The Classical method for path analysis has been used multiple linear regression for model (1) and multiple linear regression (Foreword) for three other models and summarizes the most important results through the following table:

Table (2)
Linear Regression Analysis (Classical Method)

Model		Unstandardized Coefficients		Stand. Coef.	Indirect effect	Direct effect	Std. Error	F	Sig.
		Beta	Sig.	Beta			R^2		
1	Constant	-800.52	0.030				270.53	9.417	0.000
Y	Ca	-1.281	0.093	-0.383			0.771		
	Mg	8.903	0.016	0.648	0.1437	0.4199			
	SO4	3.240	0.034	0.509	0.2358	0.2591			
	HCO3	-0.467	0.742	-0.094					
	NO3	5.608	0.215	0.273					
2	Constant	-2.304	0.956				54.423	19.21	0.000
SO4	Mg	1.550	0.000	0.719			0.516		
3	Constant	49.587	0.002				20.569	19.48	0.000
Mg	Ca	0.126	0.006	0.518	0.248	0.2683	0.696		
	HCO3	0.150	0.024	0.414	0.310	0.1714			
4	Constant	-76.359	0.132				49.747	27.79	0.000
HCO3	NO3	2.450	0.000	0.595	0.1668	0.3540	0.766		
	SO4	.590	0.002	0.462	0.2148	0.2134			

- Results and Interpretation

The path model depicted in Figure (1) Supposed, the direction of the arrows depicts the hypothesized direct and indirect paths. To estimate the magnitude of these paths, a series of regression analyses were carried out as the follows:

1. The path coefficients between (Y) and the five independent variables were obtained by regression is significant. The results from the table (1) generated from regression analysis show that Mg and SO4 entered the prediction equation (i.e., Mg and SO4 only are significant predictors). The Beta values presented in the Standardized Coefficients column represent the standardized regression coefficients between Y and Mg

equal to (0.648) as direct effect, (0.1437) as indirect effect, and SO4 equal to (0.509) as direct effect, (0.2358) as indirect effect. Mg explain 41.99% of the variation for Y, and SO4 explain 25.91% of the variation for Y, and the remainder 9.2% is explained by other variables and interactions between them.

2. The path coefficients between the SO4 and the other four independent variables were obtained by regression is significant. The results from the Coefficients table generated from foreword regression analysis show that Mg only is significant predictors and standardized regression coefficients equal to (0.719).
3. The path coefficients between the Mg and the other four independent variables were obtained by regression is significant. The results from the Coefficients table generated from foreword regression analysis show that Ca and HCO4 are significant predictors and standardized regression coefficients equal to (0.518) as direct effect and (0.248) as indirect effect, and (0.414) as direct effect and (0.31) as indirect effect respectively. Ca explain 26.83% of the variation for Y, and HCO3 explain 17.14% of the variation for Y, and the remainder 25.63% is explained by other variables and interactions between them.
4. The path coefficients between the HCO3 and the other four independent variables were obtained by regression is significant. The results from the Coefficients table generated from foreword regression analysis show that NO3 and SO4 only are significant predictors and standardized regression coefficients equal to (0.595) as direct effect and (0.1668) as indirect effect, and (0.462) as direct effect and (0.2148) as indirect effect respectively. NO3 explain 35.4% of the variation for Y, and SO4 explain 21.48% of the variation for Y, and the remainder 19.86% is explained by other variables and interactions between them.

Figure (3) presents the path model together with the estimated regression coefficients (Beta values) associated with the hypothesized paths. It can be concluded that Mg and SO4 have direct influences on the Y. The direction of the regression coefficients indicates that (1) Mg, and (2) SO4. The results also show that at least part of these influences is indirect, Where Mg affect the SO4, which in turn affect the Y, as well as our Ca and HCO3 affect the Mg, which in turn affect the Y, Finally NO3 and SO4 affect the HCO3, which in turn affect the Mg, and then affect the Y, and the following figure (Using Amos program) shows that:

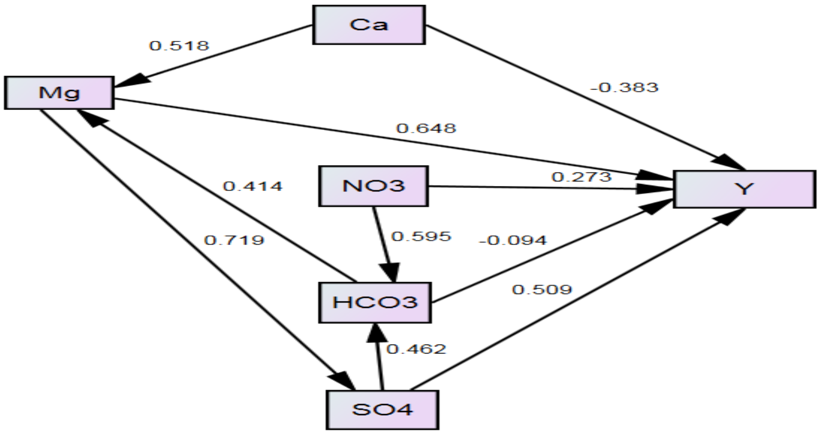

Figure (3)
The path model with the Standardized regression Coefficients by using classical method.

The correlation matrix calculation between the variables of the study (Data original) was also in the Appendix (Table-E)

3-2: Multivariate Wavelet Method:

The use of the language of MATLAB in de-noises and re-path analysis by the proposed method in Figure (1) for three multivariate wavelets as follows:

1- Multivariate Daubechies Wavelet:

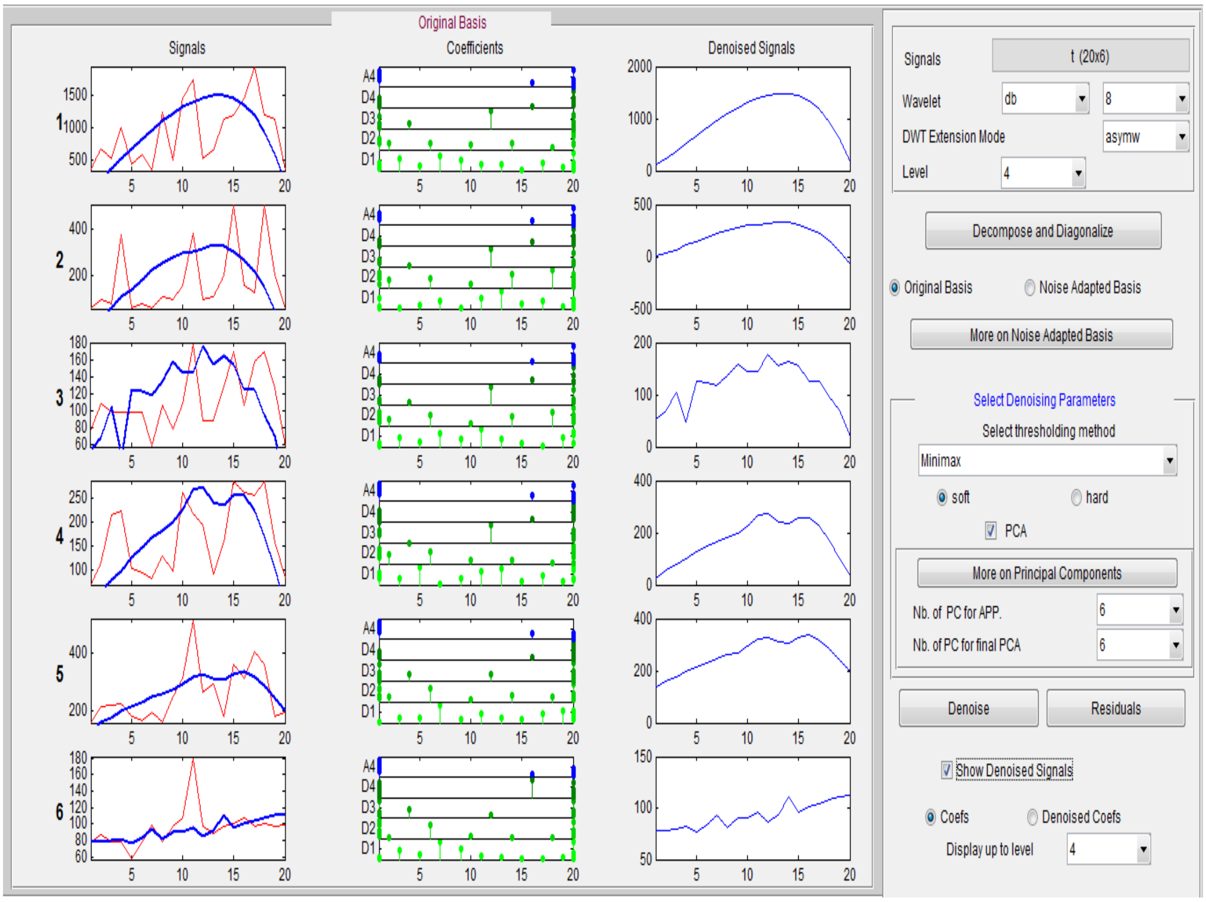

Figure (4)
de-noise by use Multivariate Daubechies Wavelet.

Check univariate distributions for normality with Kolmogorov-Smirnov test for de-noise data (see Appendix table (B)) and summarizes the most important results through the following table:

Table (3)
Kolmogorov-Smirnov Test (de-noise data-Mdb-8)

	Y	Mg	SO4	HCO3
Kolmogorov-Smirnov Z	.660	.781	.681	.624
Asymp. Sig. (2-tailed)	.776	.575	.742	.832

The results show that all p-values are greater than 0.05 (which are greater than p-values for original data) which means that all the variables have a univariate normal distribution.

Using a χ^2 test was all values of Mahalanobis distances (for multivariate normality) less than tabulated χ^2 under the significant level of 1% and 4 degrees of freedom equal to (13.28) which means that the four variables (dependent variables for four models) have a multivariate normal distribution.

De-noise data were used in the re supposed path analysis in Figure (1) and the results summarized in the following table:

Table (4)
Linear Regression Analysis (Multivariate db Wavelet Method)

Model		Unstandardized Coefficients		Stand. Coef.	Indirect effect	Direct effect	Std. Error	F	Sig.
		Beta	Sig.	Beta			R^2		
1	Constant	-417.730	0.001				32.609	784.9	0.000
Y	Ca	2.163	0.000	0.532	0.4482	0.283	0.996		
	Mg	1.054	0.056	0.098					
	SO4	.647	0.569	0.111					
	HCO3	2.137	0.087	0.280					
	NO3	1.734	0.265	0.044					
2	Constant	-19.863	0.448				39.420	61.12	0.000
SO4	Mg	1.627	0.000	0.879			0.773		
3	Constant	74.993	0.007				17.653	49.09	0.000
Mg	Ca	.416	0.000	1.107	-0.1892	1.2254	0.852		
	HCO3	-.153	0.279	-0.216	0.9697	0.0467			
4	Constant	19.313	0.211				8.285	514.1	0.000
HCO3	NO3	1.334	0.000	0.259	0.2380	0.0671	0.984		
	SO4	.680	0.000	0.891	0.0692	0.7939			

Figure (5) presents the path model together with the estimated regression coefficients (Beta values) associated with the hypothesized paths. It can be concluded that Ca has direct effect (0.532) and indirect (0.4482) on the Y. The results also show that at least part of these influences is indirect, Where Mg affect the SO4, which in turn affect the Y, as well as our Ca affect the Mg, which in turn affect the Y, Finally NO3 and SO4 affect the HCO3, which in turn affect the Mg, and then affect the Y, and the following figure (Using Amos program) shows that:

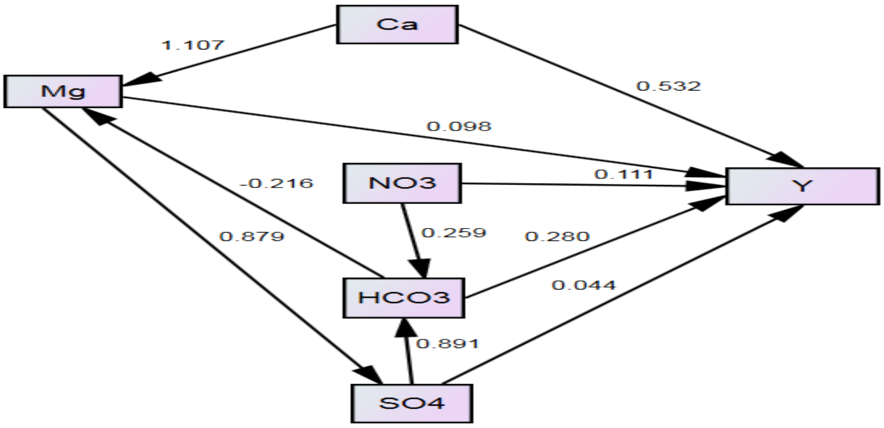

Figure (5)
The path model with the Standardized regression Coefficients by using de-noise data (Multivariate Daubechies Wavelet)

The correlation matrix calculation between the variables of the study (De-noise Data) was also in the Appendix (Table-F).

2- Multivariate bior-3.3 Wavelet:

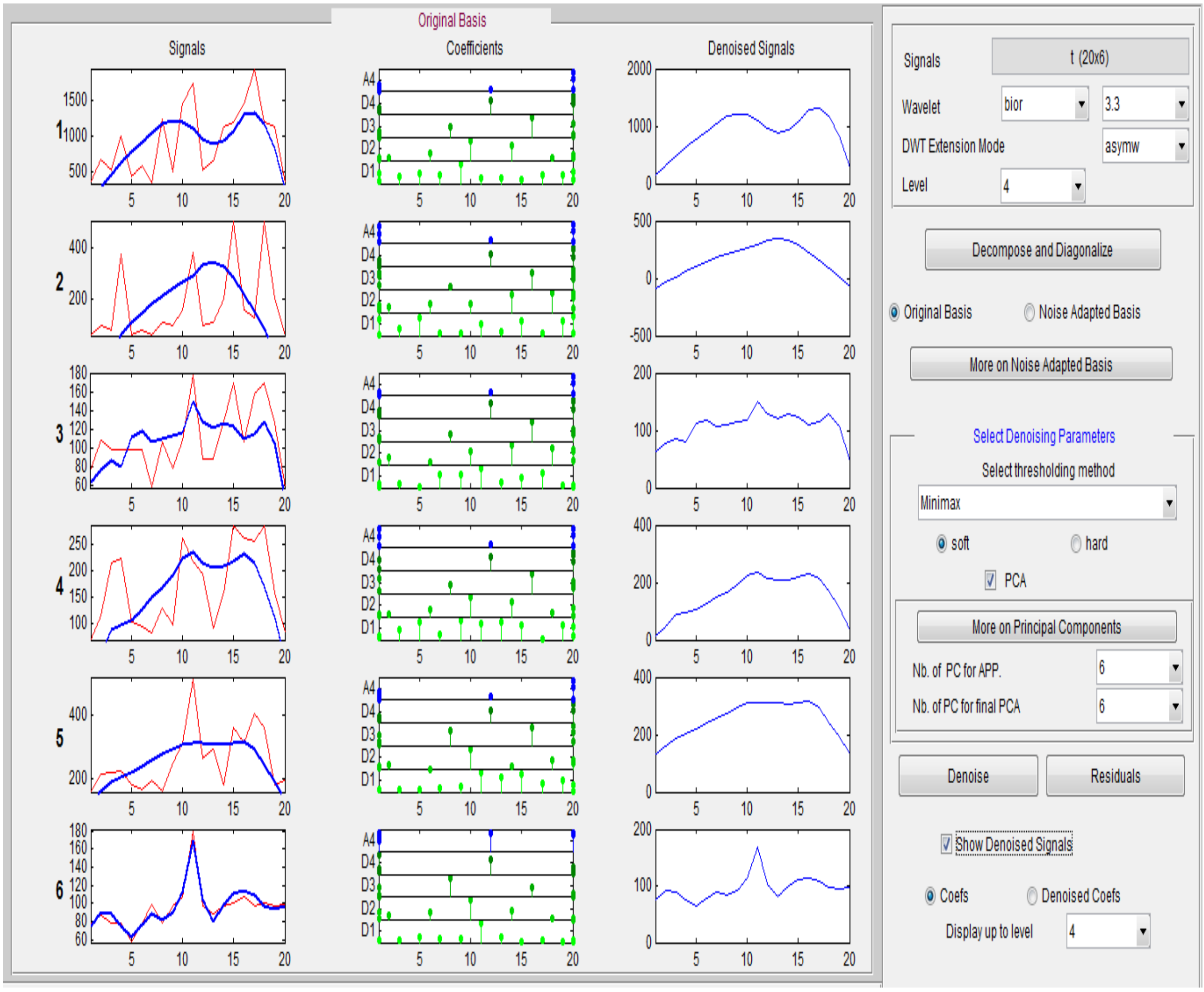

Figure (6)
de-noise by use Multivariate bior Wavelet

Check univariate distributions for normality with Kolmogorov-Smirnov test for de-noise data (see Appendix table (C)) and summarizes the most important results through the following table:

Table (5)
Kolmogorov-Smirnov Test (de-noise data- Mult. bior)

	Y	Mg	SO4	HCO3
Kolmogorov-Smirnov Z	.714	.993	.794	.874
Asymp. Sig. (2-tailed)	.688	.278	.554	.430

The results show that all p-values are greater than 0.05 which means that all the variables have a univariate normal distribution.

Using a χ^2 test was all values of Mahalanobis distances (for multivariate normality) less than tabulated χ^2 under the significant level of 1% and 4 degrees of freedom equal to (13.28) which means that the four variables (dependent variables for four models) have a multivariate normal distribution.

De-noise data were used in the re supposed path analysis and the results summarized in the following table:

Table (6)
Linear Regression Analysis (Multivariate bior Wavelet Method)

Model		Unstandardized Coefficients Beta	Unstandardized Coefficients Sig.	Stand. Coef. Beta	Indirect effect	Direct effect	Std. Error / R^2	F	Sig.
1	Constant	-932.733	0.194				112.16	34.60	0.000
Y	Ca	-2.826	0.002	-.988	1.6657	.9761	0.925		
	Mg	2.054	0.038	.144	.6619	.0207			
	SO4	1.300	0.722	.261					
	HCO3	8.203	0.101	1.478					
	NO3	-2.022	0.298	-.126					
2	Constant	-112.06	0.008				36.529	52.99	0.000
SO4	Mg	2.469	0.000	0.864			0.746		
3	Constant	14.390	0.550				13.494	23.36	0.000
Mg	Ca	-.033	0.640	-0.163			0.733		
	HCO3	.392	0.009	1.006	-0.1518	1.012			
4	Constant	139.346	0.000				9.638	402.4	0.000
HCO3	NO3	-.350	0.011	-0.121	0.6071	0.0146	0.979		
	SO4	.947	0.000	1.054	-0.0697	1.1109			

Figure (7) presents the path model together with the estimated regression coefficients (Beta values) associated with the hypothesized paths. It can be concluded that Ca and Mg have direct effects (-0.988 and 0.144) and indirect effects (1.6657 and 0.6619) on the Y. The results also show that at least part of these influences is indirect, Where Mg affect the SO4, which in turn affect the HCO3, then affect the Mg and then affect the Y, as well as our HCO3 affect the Mg, which in turn affect the Y, Finally NO3 and SO4 affect the HCO3, which in turn affect the Mg, and then affect the Y, and the following figure (Using Amos program) shows that:

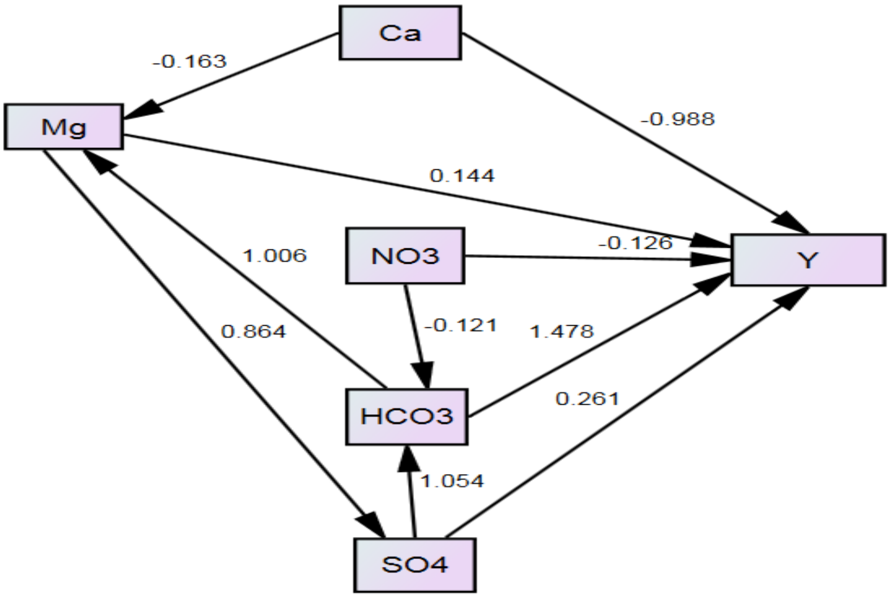

Figure (7)
The path model with the Standardized regression Coefficients by using de-noise data (Multivariate bior Wavelet)

The correlation matrix calculation between the variables of the study (De-noise Data) was also in the Appendix (Table-G).

3- Multivariate rbio-3.7 Wavelet:

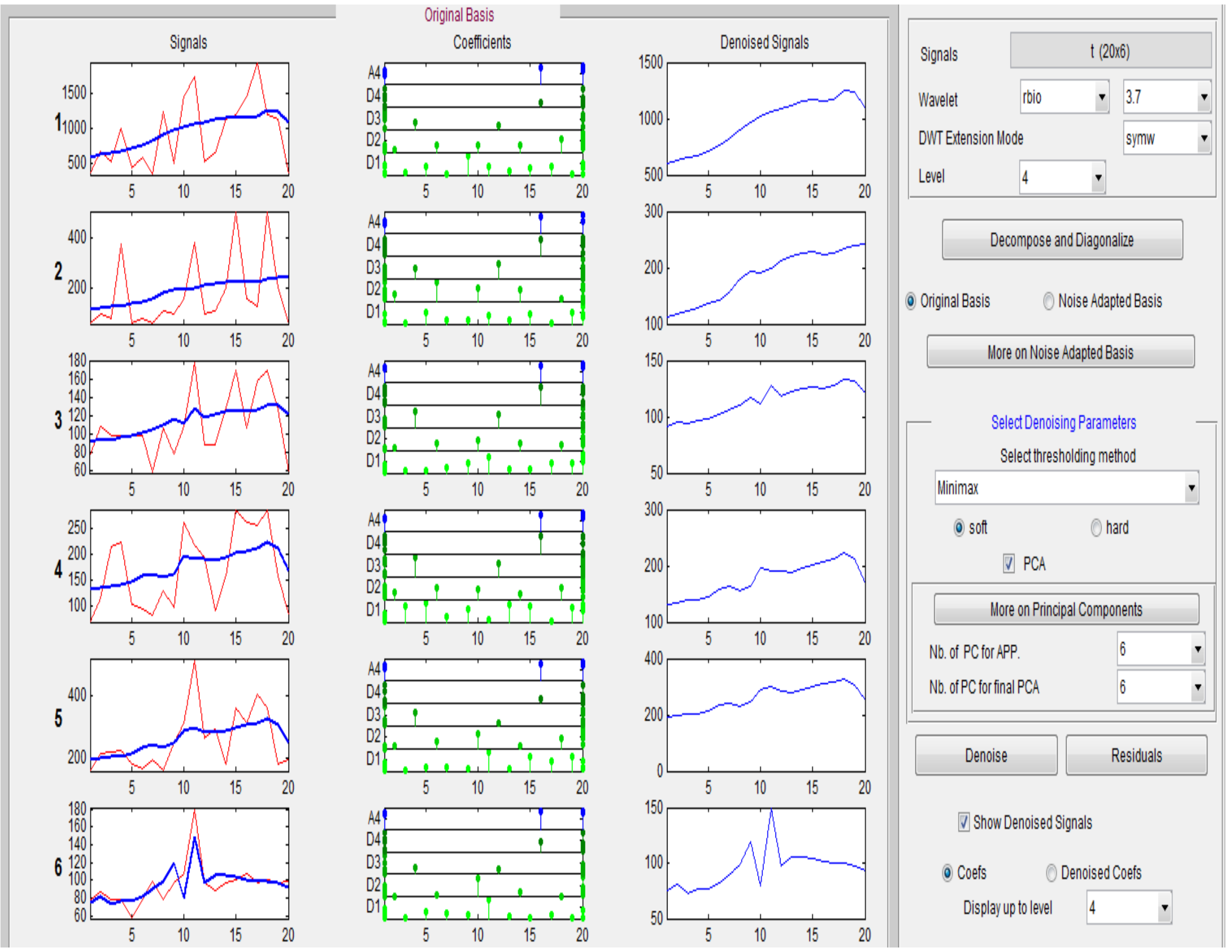

Figure (8)
de-noise by use Multivariate rbio Wavelet

Check univariate distributions for normality with Kolmogorov-Smirnov test for de-noise data (see Appendix-table (D)) and summarizes the most important results through the following table:

Table (7)
Kolmogorov-Smirnov Test (de-noise data- Mult. rbio)

	Y	Mg	SO4	HCO3
Kolmogorov-Smirnov Z	.765	.657	.788	.792
Asymp. Sig. (2-tailed)	.601	.782	.564	.557

The results show that all p-values are greater than 0.05 (which are greater than p-values for original data except SO4) which means that all the variables have a univariate normal distribution.

Using a χ^2 test was all values of Mahalanobis distances (for multivariate normality) less than tabulated χ^2 under the significant level of 1% and 4 degrees of freedom equal to (13.28) which means that the four variables (dependent variables for four models) have a multivariate normal distribution.

De-noise data were used in the re supposed path analysis and the results summarized in the following table:

Table (8)
Linear Regression Analysis (Multivariate rbio Wavelet Method)

Model		Unstandardized Coefficients		Stand. Coef.	Indirect effect	Direct effect	Std. Error	F	Sig.
		Beta	Sig.	Beta			R^2		
1	Constant	-127.671	.025				6.9059	3933.1	0.000
Y	Ca	3.420	.000	0.695	0.2885	0.483	0.989		
	Mg	-3.880	.031	-0.242	1.1959	0.0586			
	SO4	14.697	.000	1.925	-0.973	3.7056			
	HCO3	-7.695	.000	-1.514	2.4683	2.2922			
	NO3	3.287	.000	0.265	0.3422	0.0702			
2	Constant	-47.792	0.032				10.809	119.9	0.000
SO4	Mg	1.962	0.000	0.932			0.869		
3	Constant	45.740	0.000				2.782	226.7	0.000
Mg	Ca	0.181	0.000	0.590	0.3741	0.3481	0.964		
	HCO3	0.132	0.001	0.418	0.5281	0.1747			
4	Constant	-11.851	0.000				1.756	5878.7	0.000
HCO3	NO3	0.246	0.000	0.101	0.4951	0.0102	0.989		
	SO4	1.416	0.000	0.943	0.0530	0.8892			

Figure (9) presents the path model together with the estimated regression coefficients (Beta values) associated with the hypothesized paths. It can be concluded that all the independent variables have direct effects (0.695, -0.242, 1.925, -1.514 and 0.265 respectively) and indirect effects (0.2885, 1.1959, -0.973, 2.4683 and 0.3422 respectively) on the Y. The results also show that at least part of these influences is indirect, Where Mg affect the SO4, which in turn affect the Y, as well as our Ca and HCO3 affects on Mg, which in turn affect the Y, Finally NO3 and SO4 affect the HCO3, which in turn affects on Y, and the following figure (Using Amos program) shows that:

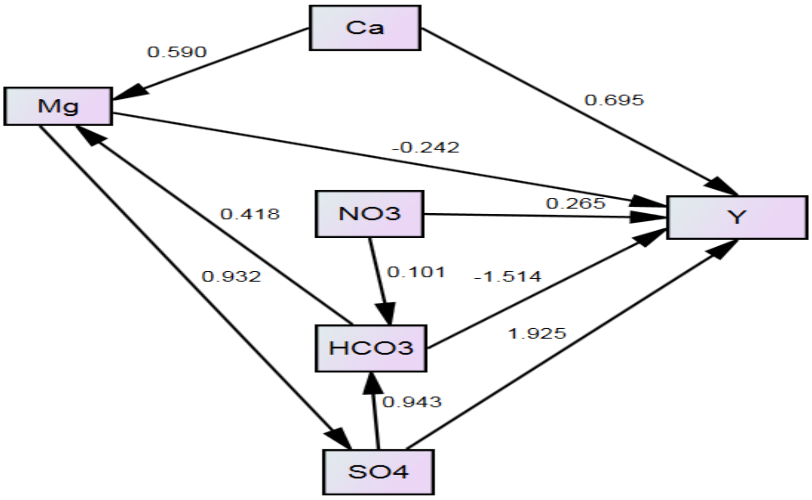

Figure (9)
The path model with the Standardized regression Coefficients by using de-noise data (Multivariate rbio Wavelet)

The correlation matrix calculation between the variables of the study (De-noise Data) was also in the Appendix (Table-H).

3-3: The comparison between the classical and proposed method

Will be here the comparison between the classical methods and proposed for a three different multivariate wavelets in the path analysis relying on some statistical criterions such as R^2, S.E. and F, and summarized in the following table:

Table (9)
The comparison between the classical and proposed method

Methods	Model	R Square	Std. Error	F
Classical	**1**	0.771	270.53	9.417
	2	0.516	54.423	19.21
	3	0.696	20.569	19.48
	4	0.766	49.747	27.79
Multivariate Daubechies	**1**	0.996	32.609	784.9
	2	0.773	39.420	61.12
	3	0.852	17.653	49.09
	4	0.984	8.285	514.1
Multivariate bior	**1**	0.925	112.16	34.60
	2	0.746	36.529	52.99
	3	0.733	13.494	23.36
	4	0.979	9.638	402.4
Multivariate rbio	**1**	0.989	6.9059	3933.1
	2	0.869	10.809	119.9
	3	0.964	2.782	226.7
	4	0.989	1.756	5878.7

The results show that Multivariate wavelets (proposed method) has best path analysis (four regression models) than classical method depending on the three statistical criterions (Because the value of S.E. is at least while the R^2 and F are the largest), So was the way Multivariate rbio wavelets (proposed method) is the best compared with other the methods used in the path analysis (four regression models) depending on the three statistical criterions.

4: Conclusions and Recommendations

Through this study we reached to the following Conclusions and recommendations.

4-1: Conclusions:

1. The proposed method is better than the classical method for Path analysis of these data.
2. The proposed Multivariate rbio method is better than the other methods for Path analysis depending on the statistical criteria used.
3. There is amount the effect of the independent variables (direct and indirect) different on dependent variable depending on the method used in the treatment of contamination.

4-2: Recommendations:

1. Use the proposed methods in the path analysis, especially Multivariate Wavelet rbio for other data.
2. The proposed method compared with other methods for contamination treatment.
3. Conduct a study on the use of the method proposed in Structural Equation Modeling.

5: References:

[1] Alwin, D. F., & Hauser, R. M. (1975). The decomposition of effects in path analysis. *American sociological review*, *40*, 37-47.

[2] Andreas de Vries, (2006), Wavelets, FH Sudwestfalen University of Applied Sciences, Haldener Straße 182, D-58095 Hagen, Germany.

[3] Donoho, D.L.; Johnstone, I.M. Adapting to Unknown Smoothness via Wavelet Shrinkage. J. Am. Stat. Assoc. 1995, 90, 1200–1224.

[4] Kaijian He, Kin Keung Lai, and Guocheng Xiang, (2012), Portfolio Value at Risk Estimate for Crude Oil Markets: A Multivariate Wavelet De-noising Approach, Energies,5 , 1018-1043.

[5] Loehlin, J. C., & Beaujean, A. A. (2017). Latent variable models: An introduction to factor, path, and structural equation analysis (5thed.). New York, NY: Routledge.

[6] Rex B. Kline (2004). Principles and Practice of Structural Equation Modeling 2nd Edition, New York: Guillford Press.

[7] Simeonov, V., Stratis, J.A., Samara, C., Zachariadis, G., (2003) Assessment of the surface water quality in Northern Greece. *Water Research*, **37**, 4119-4124. doi:10.1016/S0043-1354(03)00398-1

[8] Wang, E. Mandell and M. Rose, (2006) Digital dark field tableaus, Microscopy and Microanalysis 12:S2, 1010–1011.

(1) Appendix

Table (A): Original data

Y	Ca	Mg	SO4	HCO3	NO3
365.44	58.22	78.25	69.26	159.26	78.22
671.96	97.66	107.69	114.83	214.33	87.66
519.29	78.30	98.90	215.52	215.72	78.30
1003.87	377.56	97.96	223.56	223.76	77.56
429.70	57.52	97.82	102.21	176.23	57.62
576.19	78.43	98.93	94.82	164.62	78.73
345.95	58.24	58.24	81.21	191.11	98.28
1228.24	107.11	107.11	127.77	157.77	77.61
501.95	97.96	77.96	97.07	247.07	97.96
1451.51	157.98	107.88	260.88	310.48	107.78
1731.95	379.95	179.15	217.41	517.41	179.95
532.49	97.80	87.87	192.95	262.85	97.80
654.17	107.88	87.98	89.74	294.70	87.98
1120.36	197.99	127.99	158.51	178.61	97.79
1181.18	499.40	169.40	283.47	363.17	99.90
1451.51	157.99	106.80	260.81	310.44	107.71
1910.79	127.95	157.85	255.32	405.14	97.15
1180.19	500.40	169.50	283.57	363.18	100.10
1121.32	196.99	126.99	158.52	178.62	97.74
342.95	58.14	58.34	81.31	191.13	98.22

Table (B): De-noise data (by using multivariate db wavelet)

Y	Ca	Mg	SO4	HCO3	NO3
129.19	3.24	53.75	26.73	136.44	78.14
241.35	30.00	68.56	55.14	158.19	78.72
369.38	60.98	105.05	79.03	174.49	79.62
516.74	110.10	48.53	97.57	197.66	81.90
664.98	140.38	124.94	125.80	212.74	77.26
817.07	182.19	123.82	147.01	230.13	83.70
961.73	221.26	117.75	167.34	246.56	93.75
1092.77	254.26	135.14	181.70	259.51	81.24
1207.16	278.86	157.90	199.26	270.15	90.02
1307.92	298.14	144.57	226.89	292.36	90.23
1391.00	305.06	145.85	265.81	319.19	96.11
1449.58	312.58	177.03	272.33	324.74	85.86
1482.66	332.07	155.50	240.80	309.89	92.87
1479.73	326.77	164.49	235.26	306.72	111.63
1435.80	300.35	154.63	255.59	327.73	95.75
1341.98	266.71	125.17	256.55	335.22	101.33
1182.18	219.44	125.47	223.09	316.83	103.61
937.77	151.00	94.79	170.25	286.37	107.90
597.08	50.46	69.95	102.59	242.10	111.00
	00.	19.80	36.01	198.77	111.72

Table (C): De-noise data (by using multivariate bior wavelet)

Y	Ca	Mg	SO4	HCO3	NO3
140.13	00.	64.27	13.59	134.61	75.71
304.42	00.	77.19	46.75	157.59	90.87
466.79	11.13	86.15	86.61	185.96	89.09
628.16	62.08	80.66	96.96	204.80	74.87
777.56	107.92	111.07	104.99	218.91	63.38
917.19	146.18	118.43	125.86	237.96	77.14
1046.91	180.14	106.93	149.21	258.53	87.80
1165.45	211.26	110.69	166.29	275.04	82.35
1215.09	239.07	113.99	190.69	292.00	90.95
1195.85	263.67	116.30	222.27	309.44	112.33
1106.86	291.36	150.87	235.78	313.21	168.22
951.19	332.09	129.23	215.34	310.00	103.51
891.15	344.11	121.46	206.47	308.26	80.42
926.76	327.45	127.57	209.17	308.00	98.97
1058.82	285.27	123.55	217.76	310.76	110.37
1287.35	217.57	109.40	232.24	316.53	114.63
1322.39	150.62	115.30	215.26	293.29	109.02
1164.42	83.93	128.25	169.96	243.14	97.48
814.58	10.44	105.45	110.07	190.31	93.33
272.88	**00.**	**46.90**	**35.60**	**134.80**	**96.56**

Table (D): De-noise data (by using multivariate rbio wavelet)

Y	**Ca**	**Mg**	**SO4**	**HCO3**	**NO3**
595.91	113.20	91.69	131.23	191.83	74.85
625.19	118.89	94.59	134.34	197.87	81.47
650.21	123.82	94.23	139.08	202.85	73.06
673.08	129.25	96.04	139.40	204.12	76.19
710.84	136.09	97.93	145.46	213.29	76.67
759.38	142.28	101.43	158.11	233.65	81.90
824.20	156.54	105.59	162.34	241.21	89.59
904.42	180.02	110.14	155.60	231.65	98.97
969.60	192.05	116.99	162.67	246.68	118.61
1018.23	191.63	111.92	196.30	288.82	79.45
1058.84	199.34	128.31	191.33	297.68	149.05
1087.79	212.30	118.78	190.83	282.05	96.87
1115.04	220.21	122.08	188.88	280.20	105.25
1149.32	225.91	124.35	195.01	288.98	105.58
1165.50	227.31	125.74	201.22	298.26	103.96
1150.06	222.94	125.60	206.73	308.13	101.31
1170.85	225.17	127.27	211.29	313.72	99.55
1254.63	234.25	132.54	222.27	324.83	99.40
1236.30	239.53	131.19	210.94	307.34	97.64
1087.77	241.33	121.00	168.62	248.75	92.62

Table-E: Correlation Matrix (Data original)

		Y	Ca	Mg	SO4	HCO3	NO3
Correlation	Y	1.000	.528	.789	.742	.698	.548
	Ca	.528	1.000	.766	.699	.599	.420
	Mg	.789	.766	1.000	.719	.724	.499
	SO4	.742	.699	.719	1.000	.677	.361
	HCO3	.698	.599	.724	.677	1.000	.762
	NO3	.548	.420	.499	.361	.762	1.000

Table-F: Correlation Matrix (De-noise Data)

		Y	Ca	Mg	SO4	HCO3	NO3
Correlation	Y	1.000	.980	.893	.988	.948	.289
	Ca	.980	1.000	.917	.960	.876	.147
	Mg	.893	.917	1.000	.879	.754	-.026-
	SO4	.988	.960	.879	1.000	.960	.267
	HCO3	.948	.876	.754	.960	1.000	.497
	NO3	.289	.147	-.026-	.267	.497	1.000

Table-G: Correlation Matrix (De-noise Data)

		Y	Ca	Mg	SO4	HCO3	NO3
Correlation	Y	1.000	.678	.807	.899	.877	.409
	Ca	.678	1.000	.774	.880	.931	.408
	Mg	.807	.774	1.000	.864	.854	.485
	SO4	.899	.880	.864	1.000	.985	.576
	HCO3	.877	.931	.854	.985	1.000	.486

	NO3	.409		.408	.485	.576	.486	1.000

Table-H: Correlation Matrix (De-noise Data)

		Y	Ca	Mg	SO4	HCO3	NO3
Correlation	Y	1.000	.985	.983	.954	.954	.608
	Ca	.985	1.000	.964	.894	.895	.588
	Mg	.983	.964	1.000	.932	.946	.722
	SO4	.954	.894	.932	1.000	.996	.525
	HCO3	.954	.895	.946	.996	1.000	.596
	NO3	.608	.588	.722	.525	.596	1.000