MATHEMATICAL MODELING METHODOLOGY FOR QUALITY PARAMETERS OF SURFACE WATERS BODIES

ANTOHE VALERIAN[1*], STANCIU CONSTANTIN[1]

[1] *"Dunarea de Jos" University of Galaţi, Faculty of Engineering Brăila, Calea Călăraşilor, No. 29, RO-810017, Brăila, Romania*

Abstract: Surface waters are the main source of water consumption but at the same time, the majority waste waters are discharged in the surface water bodies. In order to highlight the evolution of environmental quality parameters and especially the evolution of this process and to achieve a set of procedures for improvement of some necessary steps to optimize the system of observation and action in time we propose this study. In the last three years a study has focused on the drawing of a rich map of data for the quality of surface water of the Danube River was developed and the main idea was to consider a great set of data and to organize them in a mathematical model in order to have an exhaustive image of the evolution of the water quality in time. Modelling this database using specific results which belong to numerical mathematical analysis was a desire of ours and the results can be used now in order to present an evolution, to make prediction and to consider the Danube River like a human entity which lives following mathematical rules that could be modelled.

Keywords: mathematical modelling, quality parameters, art of analyzing and optimization.

1. INTRODUCTION

In this study we followed an analysis of the Danube River water quality during the years 1992-2006 for the main body of surface water, the Danube River, considering a distance on the Danube River between km 132 and km 375. A mathematical model was designed for analyzing the specific data with methods of numerical analysis, focused on time series of data, using different methods of mathematical analysis performed in different programming environments. The development of this new mathematical model will reflect in a more realistic way the impact of major urban areas in the studied area (the towns of Braila and Galati) and their influence on the Danube water quality and the importance of wetlands on the level of surface water bodies. Specialized studies cover a wide range of theoretical models of calculating the evolution of water quality parameters of surface waters but the problem of predicting the evolution of these parameters has always been considered a difficult one. Thus, this study is trying to consider all the sets of quality parameters for surface water. The analysis covers step by step all the subset of the quality parameters. Our aim is to study a particular segment of an area and this means that the model must be adapted to the situ, accepting specific conditions to be studied, taking into account the particularities for the mathematical model parameters embedded in the specific area under consideration. Another condition is to consider in a mathematical model a simplified model which covers only one group of quality parameters. This method of simplified models is one accepted in different studies and the connections of these subsystems could cover the evolution of the complexity of a model for water environment.

* Corresponding author, email: valerian_antohe@yahoo.com

The present mathematical model describes the quality of surface water body Danube, covers the category of dissolved oxygen and its group of quality parameters, and shows the dynamics of this group, while also proposing estimates based on data analysis of a beach rich in time.

Mathematical model design took into account the mathematical study of mathematical statistics, the numerical analysis and the use of computing platforms, such as MATLAB, [1], or Table Curve 2D. For the current study were considered theoretical notions aiming time series, [2], numerical analysis, [3,4], and other works in the field aiming to shaping a time series especially in terms of modeling a time series and achieving a time of prediction [5]. The literature shows the predictability of time series, but with significant results and impact in practice for economic or geo-demographic statistics, [6], and less with significant results for the prediction of environmental phenomena on the quality of water bodies and their evolution for a long-time in the future. In the model construction, some theoretical results are established. The objectives stated in Management Plan of the Danube River Basin, (PNMBD) are focused on the protection, preservation and restoration of wetlands and floodplains, to ensure biodiversity, flood protection and pollution reduction by 2015. Necessary connection with water temperature, levels or rain flow and water flow will be considered in the future. The present paper will following the next structure: consideration about the structure of the database and its mathematical theoretical and practical organisation, the structure of the visible space of observation and prediction, exploration of the best continuous function which could modelling the real evolution of a quality parameters and some open problems for this study.

2. DATA-BASE, TIME SERIES OF QUALITY PARAMETERS OF SURFACE WATERS

2.1. Data-base. State function of a quality parameter *C*

Considering the studied area, we have used quality data collected over time in four monitoring stations on the Danube as follows in Table 1, from Trans-National Monitoring Network, (TNMN), of International Commission for the Protection of the Danube River, (ICPDR) and National Agency Romanian Waters, (ANAR).

Table 1. Monitoring station considered for the study.

No	Monitoring station	Indicative	Longitude	Latitude	Km on Danube	Altitude (m)	Station indicator
1	Chiciu/Silistra	L0280	44.121667	27.243889	375	13	S1
2	Gropeni	Gr	45.08	27.90	180	5	S2
3	Braila 1	Br 1	45.18	27.93	176	4	S3
4	Reni-Chilia/Kilia	L0430	45.482216	28.226624	132	4	S4

Definition 1. A ***time series*** is the sequence of data (usually finite), representing the observations taken at time t_i for a random variable X_i. Time series are finite strings of real numbers that represent the result of observations of phenomena in various fields. In this case, the result is the observation of one quality parameter of surface water.

Table 2. Time series for (DO), Station S1, 1999. There are no recorded data for October and December.

Time Monthly	1	2	3	4	5	6	7	8	9	10	11	12
The concentration of (DO) in mg/L	11	10.7	9.8	9.3	6.9	6.7	6.4	7.2	6.7	-----	8.32	----

One could see that for 1999, there are no registered data for two months, namely October and December. The question could be, how should be indicated a level of dissolved oxygen, (DO) concentration for these months with no registered data? The mathematical construction that will be developed here, will give some possible answers. The idea is to use quadratic Spline function, considering the data from Table 2 and the data registered in January 2000 and in this case data will be obtained for October and December 1999. In this manner the result for missed data will be: 7.2732 mg/L for October and 9.3705 mg/L for December. In order to cover all other missing data we must propose the model for all registered data between 1992 and 2005.

Proposition 1. Each quality parameter posses a time series, parameter values as specified for which data were recorded after achieving a monitoring process and the data are continuous in time, even if these are not the result of a process of standard analysis in a specific laboratory.

The proposition 1 opens the route for the current investigation and allows us to consider that data missing could be modeled using different mathematical methods. Investigating the evolution of the quality parameters with proposed method which uses Spline function one could see that our modeling appropriate well the evolution. Such a result is realized in the data for probe 2 showed in table 3 in the 8-th month in the same meaner used for covering the missed data in table 2 for the 10-th and 12-th month. Data covered are in the brackets in Table 3 and all the values belong to the modeled data for year 2005 (2 probes registered for every month), considering (DO) in S1.

Table 3. Concentration of (DO), two probes per month, P1, P 2, monthly, year 2005, station S1.

Station 1 L0280 2005 Dissolved Oxygen concentration in mg/L												
Month	1	2	3	4	5	6	7	8	9	10	11	12
Probe1	12	11.2	10.8	8.2	7.7	6.3	5.6	6.4	5.2	5.5	8.2	8.9
Probe2	11.7	11.5	10.1	9.7	5.8	10.2	7.1	**(5.8)**	5.1	5.7	6.8	8.6

Definition 2. String values registered during a period of time for any quality parameter will be named ***temporal series of a quality parameter C*** and will be noted in further with, $\tilde{X}_C$.

Definition 3. For a period of time [0, T], it will be called ***state function of a quality parameter C***, a function X_C: $[0,T] \to R$, that will have the following properties:

- The restriction of the function X_C over the set $\{t_0, t_1, ..., t_k\} \subset [0,T]$ will be the temporal series of a quality parameter C;
- The proposed function is a bordered one, on the interval [0,T];
- The function X_C is a continuous and differentiable one.

Some observation must be considered, with respect to these definitions.

Observation 1. The timing "0" of the observation for the interval [0, T], may be the most distant point in past time for which observable statistical data exist and are considered, (for example, the year 1992).

Observation 2. Time series are, for a quality parameter of the string type in general, meaning as functions of real numbers defined on a set of natural numbers, corresponding to the results of observations on a phenomenon at different times t_0, t_1, t_2,... so the real numbers of string type $\{X_C(t_i)\}_{i=1,n} \subseteq R$ corresponding to moments where a given quality parameter was recorded, (daily, monthly, bimonthly, etc.). Definition 3 extends the time series characterized by discrete values to a function defined on a compact interval [0, T], and therefore leads to the following working hypothesis: for all the quality parameters of water bodies, the time evolution of a particular parameter behaves in reality as a continuous and differentiable function throughout the interval [0, T], this function will initially be known only by discrete values corresponding to values recorded in samples taken monthly, bimonthly or on an annual basis.

Observation 3. The function XC: $[0, T] \to R$, which associates the value of XC (t) $\in$R, $\forall$ t$\in$ [0, T], to a quality parameter over the set {t0, t1, ..., tk}, in which |ti+1-ti|=constant, $\forall$ i $\in$ N, we will note the further with $\tilde{X}_C$.

In practice, this function corresponds to the temporal series of quality parameters, if and only if the quality samples were performed on the equal time intervals. This is rarely possible but the values considered in the analysis database will be used as a starting base in order to build from them, the continuous functions.

For example, for Dissolved Oxygen, (DO), the function $\tilde{X}_C : \{1,2,3,...,12\} \to R$ could be represented as follows in Table 4, where the data sets for 12 probes per year 1996 are presented.

Observation 4. In the present paper the numerical examples cover different years in different presented data table in order to present more realized data for interval of time 1992-2005. The same construction is realized in a large data table for all the 15 considered years.

Table 4. Temporal series of a quality parameter ***C*** = (DO), the year 1996.

(DO)	**L0280**		**mg/l**								**C≡**	**(DO)**
i	1	2	3	4	5	6	7	8	9	10	11	12
t_i	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}
$\tilde{X}_C$ (t_i)	7.9	8.04	8.25	7.5	6.39	5.7	7.42	5.9	7.48	8.4	8.87	9

2.2. Visible space of the quality parameters (VS)
Type series are plotted as follows in the report of the majority of agencies involved in monitoring parameters of environment quality, (Figure 1):

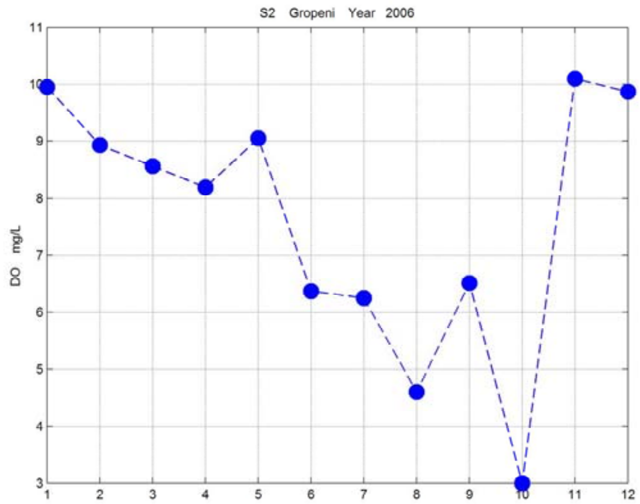


Fig. 1. $\tilde{X}_C$ Series map for ***C***= (DO), year 2006, station (S2), Gropeni.

In case of sample collection at long intervals of time, the presence of some values out than the permissible limit concentration could be hidden for observation, and thus, any measures to counteract the phenomenon of risk could be delayed or impossible to control. In case of data with the help of the interpolating with cubic spline function on a very fine mesh interval (1:1000) one could see the parameter evolution, much closer to real physico-chemical variation of dissolved oxygen concentration, maybe, with some evolutions which exceed the minimum and maximum values achieved in the first analysis of data (Figure 2).

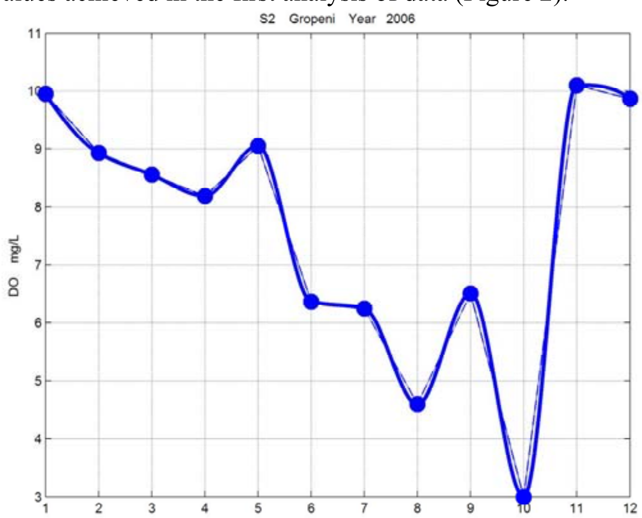


Fig. 2. X_C series map for ***C***=(DO), year 2006, station (S2), Gropeni interpolated with cubic Spline.

In order to connect all these maps for many parameters of water quality, we will define the ***visible space of the quality parameters*** **(VS).**

Quality parameters of surface water bodies and water bodies in general are presented with its units of measuring. The comparison of these data and carrying out investigations on the interactions between different chemical in aquatic space species requires bringing all these values in a common space of observations and finding common traits in the analysis of their evolution in time. For this reason we introduce the next definition:

Definition 4. The ***visible space of the quality parameters,*** **(VS),** will be considered a domain which belongs to R^2, of the form: (VS)=[0, T]x[0, M], where T,M∈R_+, with these properties:

- For every parameter of quality C, the max and min values of the function X_C belong to (VS);
- The extension of every function X_C on the interval of the form [T, T_1] included in the interval [0, ∞) has the extreme values included in the interval [0, M];
- There is a finite number of functions X_C which could be null almost everywhere in the domain (VS), (Figure 3).

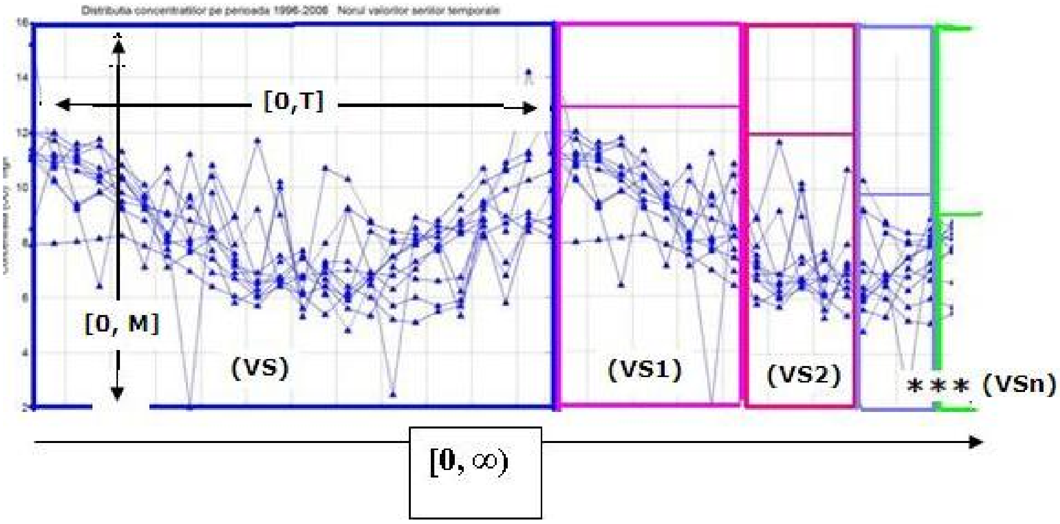


Fig. 3. Visible space of the quality parameters, (VS).

Definition 5. The domain [T, T_1]x[0, M] will be named the visible space of predictions and it will be noted as (VSi).

Proposition 2. Every (VS) possesses a space (Vsi) and the reunion of these two disjoint spaces will form another space of type (VS).

Observation 4. Proposition 2 will generate iteratively, a sequence of spaces of the form (VSi), so, one could construct a string of spaces of the next prediction $\{VS_i\}_{i=1,n}$. The (VSi)=[T_i, T_{i+1}]x[0, M] domains are characterized by its norm as absolute value of the difference T_{i+1}-T_i. Otherwise, the string $\{VS_i\}_{i=1,n}$ depends on the (VS) domain and methodology used to realize the prediction considering the trend of the evolution of the quality parameters values.

Proposition 3. The string of norms of the domains (VSi) tends to zero.

2.3. Standardization of time series using cubic Spline function

Database for water quality parameters, used by different agencies, include samplings of 6, 12 or maximum 24 data per year. In order to obtain uniformity for all the data considered for a year, cubic Spline interpolations were used. Thus, uniformity was obtained for all parameters, realizing 24 data per one year. For example, for (DO) concentration registered in S1 station containing 12 data, Table 5 was processed using a MATLAB description with Spline interpolation and data were upgraded to 24 data per year, 1992. Data used in the model will be in conforming to Table 6.

Table 5. Initial data set for (DO) concentration in mg/L, monitories in 1992 registered in (TNMN) archive.

Month	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
DO In mg/L	12.2	7.7	10.8	8.3	7.5	6.5	7.3	6.4	6.2	6.9	10.2	9.8

Table 6. Data set obtained after an interpolation with cubic Spline function was applied to data belongs to table 5.

Nr.Crt	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
DO	12.2	***8.9125***	7.7	***9.25***	10.8	***9.7015***	8.3	***7.8596***	7.5	***6.8889***	6.5	***6.9***

Nr.Crt.	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
DO	7.3	***6.8909***	6.4	***6.2591***	6.2	***6.4056***	6.9	***8.6944***	10.2	***10***	9.8	***9.8***

In this way we have standardized all data to be analyzed, and graphical analysis shows that one can observe any trend that describes the variation of dissolved oxygen concentration. Analyzing different situations of such developments one can show critical periods that are consistent with what was given by the drought and high temperatures as well as some cases explained by a greater level of rainfall, [7].

The processing of statistical data with methods of interpolation for the period of 15 years monitoring, for example for (DO), produces a time evolution which becomes more complex and may involve more extensive analysis. Such graphic images of the temporal evolution of a particular quality parameter could be made for any point of monitoring, where there is a historical monitoring data. For example, we considered the evolution of (DO) using 24 data per year (obtained with cubic Spline interpolation) multiplied by 15 years of observation and a detailed image of the evolution map was obtained, (Figure 4). In the same picture one could observe the linear regression function processed by MATLAB, of the form y=-0.00018x+0.6, expressing a descending linear trend. On the xox' axes each division represents one year and on the yoy' axes one could see the level of (DO) in mg/L.

The construction drawn produced a data base consisting of 360 data for all the parameters of quality of water for 15 years, from 1992 to 2006. If graphic functions overlap for two parameters, there can be made connections, showing the influence and interconnections between different water quality parameters.

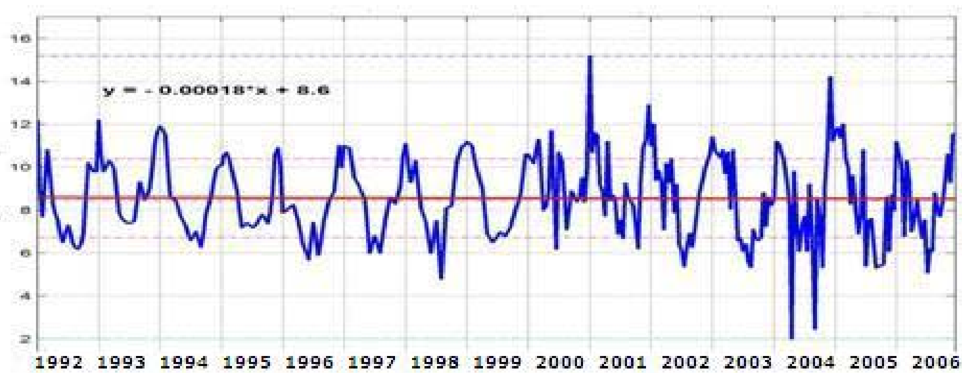


Fig. 4. Evolution of the concentration of (DO) according to the database in the monitoring station S1, considering all 15 years x 24 data.

Data considered could cover all the data for one month for all 15 years and, in this case, one can analyze the evolution of the parameter of quality during one specified month. This will show an evolution specific to a period of one month considering a rich database. In Figure 5, was considered as parameter of quality the bio-chemical consumption of (DO) at five days, (CBO5) with two data for every year in May, for two stations S1 and S4. Analyzing the map, one could conclude that there are some critical periods in May, when it is registered a possible pollution (period between 1992 and 1996) on the segment of the Danube River considered as analyzed situ. On the xox' axes, every two divisions from all 30 divisions represent one year.

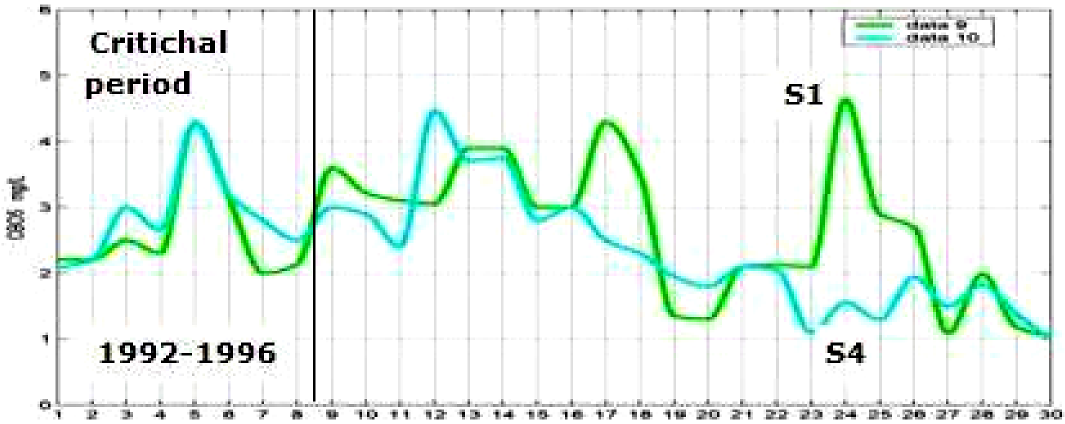


Fig. 5. Data for CBO5, with all the data obtained with Spline interpolation, considering all the data for one month, May in every year 1992-2006.

3. MATHEMATICAL MODEL ON (VS)

The considered model used for the mathematical construction contains the set of real function F, including the set of continuous and differentiable function C[0, T], and the set of polynomial function $\{P_n\}$ used for the interpolation of temporal series $\tilde{X}_C$ of a quality parameter C and the set of state function X_C, (Figure 6).

The most important problem is to close the diagram between $\tilde{X}_C$ and X_C.

In figure 6, ***X*** represents the set of temporal series for any parameter of quality considered for a parameter of quality ***C***, ***C [0, T]*** is the set of continuous functions over the interval [0, T], which could be the representation of a state function of a quality parameter C and is the set of polynomial function which could interpolate every temporal series. For any temporal series $\tilde{X}_C$ there is a string of polynomial function which could interpolate the temporal series $\tilde{X}_C$ and this string approximate uniform and converge uniformly to the function X_C.

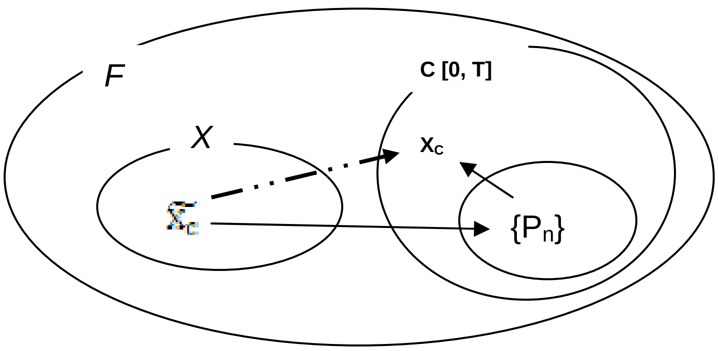
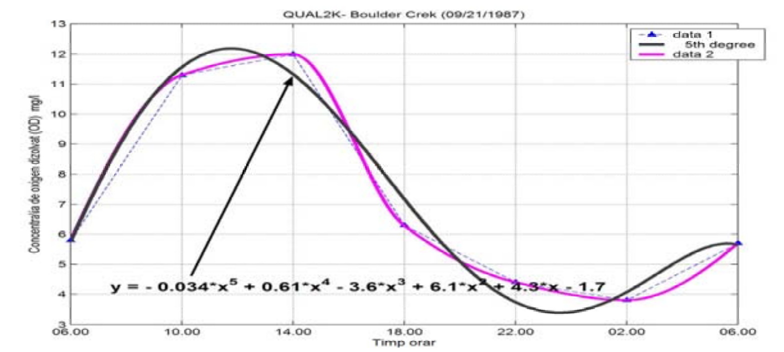


Fig. 6. The construction of the state function X_C requires the closure of the map.

In this way we are building in the context of the Weierstrass theorem but in another sense. In fact, the Weierstrass theorem affirms that every continuous function is the uniform limit of a string of polynomial functions, „id. est.", a function X_C, can be approximated by a string of polynomial functions. In the construction we have not an analytical form of the X_C function. A lot of contexts are used in order to appropriate this function. For the construction, we have considered only some values of the functions $X_{C,}$ we have used some methods of interpolation using polynomial function but, the uniform limit of the string of used polynomial function will not help us to find the analytical form of the function X_C.

3.1. Polynomial functions which converges uniformly to X_C

In fact, investigating a lot of data of quality for the constructed data base, we will manipulate very well the values of the function X_C, and sometimes, the polynomial form is very appropriate of the function X_C, in the sense of manipulating values of X_C. The Weierstrass theorem considers the continuous function X_C, and the string of polynomial functions which converges uniformly to X_C, in the same context as the set of real numbers R and the set of rational numbers Q. This will give the route of finding the best polynomial form which approximates very close to reality a state function X_C, for one quality parameters. Such an example is shown in fig. 7, and another in fig. 8, accepting that other such examples were studied. In figure 7 and 8 data were studied for (DO) performed with the QUALE2 program in order to use some data as these appear in an evolution with data processed every two hours. For all these investigation with polynomial forms, we must accept that the polynomial form will model well the evolution if and only if the residual norm will be the smallest.



$y = p1*x^5 + p2*x^4 + p3*x^3 + p4*x^2 + p5*x + p6$

Coefficients in polynomial form

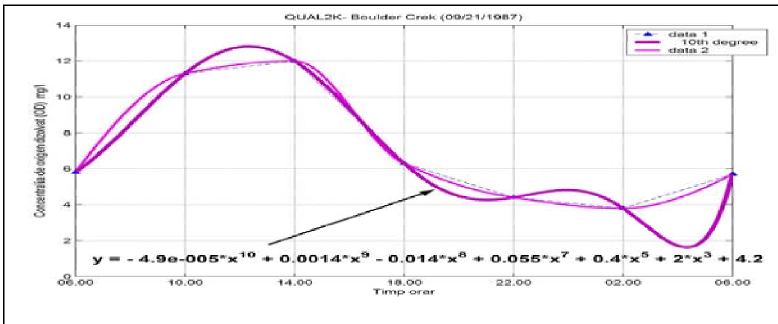
p1 = -0.03375
p2 = 0.61061
p3 = -3.5593
p4 = 6.1083
p5 = 4.287
p6 = -1.6571

Norm of residual = 1.3455

Fig. 7. Seven data were analyzed and the polynomial of degree 5 approximate very well the evolution.

For a 24 per year data, a polynomial form of 7-th degree has the form (Equation 1), and models with a fine residual form the evolution of the (DO) in mg/L, during one year.

$$y = 7.1e\text{-}007x^7 - 6.7e\text{-}005x^6 + 0.0026x^5 - 0.051x^4 + 0.56x^3 - 3.4x^2 + 9.3x + 3.6 \quad (1)$$



$y = p1*x^{10} + p2*x^9 + p3*x^8 + p4*x^7 + p5*x^6 + p6*x^5 + p7*x^4 + p8*x^3 + p9*x^2 + p10*x + p11$

Coefficients of the polynomial form

p1 = -4.9405e-005 p2 = 0.0013975 p3 = -0.014146 p4 = 0.054874
p5 = 0 p6 = -0.40218 p7 = 0 p8 = 1.9918 p9 = 0 p10 = 0
p11 = 4.1683

Norm of residual = 6.7718e-012

Fig. 8. The degree of polynomial form increase shows an anomaly of the evolution.

Many polynomial forms could be used in order to show the evolution of a parameter. The polynomial form is specific to one defined situation. Connecting the polynomial forms, continuum one could find the analytical form of X_C state function, (Figure 9).

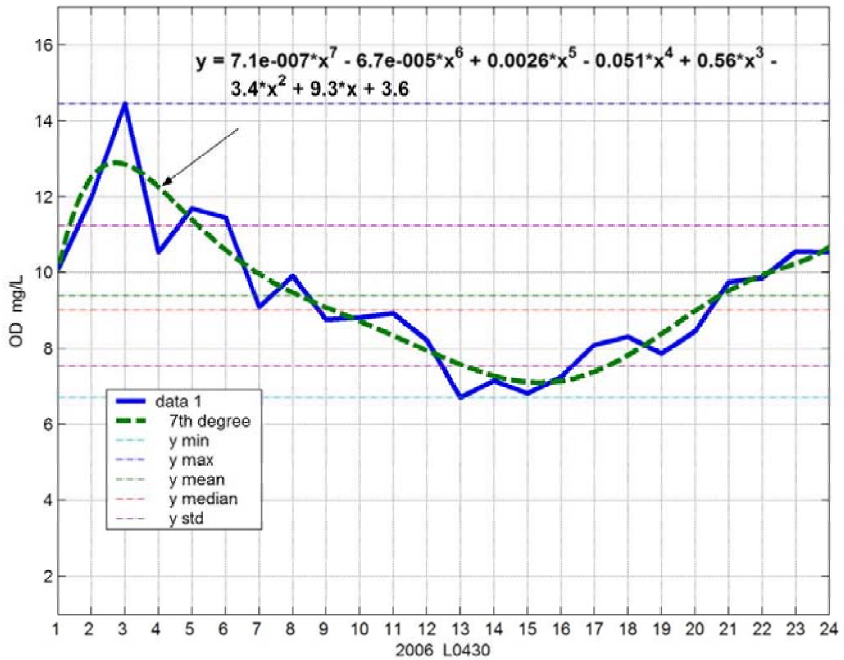


Fig. 9. Real data, interpolated and the evolution is very well modelled by a polynomial form of 7-th degree.

3.2. The best approximation for evolution and prediction
When an evolution it is analyzed for one year with 24 registered data, the polynomial forms could show a trend and an evolution with an accepted residual norm. If an analysis covers a long term evolution, (15 years), any polynomial form will not cover a visible evolution which could show for example the difference of the level of (DO) registered in the first S1 station (entrance point for us), and the level of (DO) for registered data in the last station S4, after the Danube River cross Braila and Galati city zone.

For this reason, the construction has followed some necessary steps. First of all a mesh with more than 3500 data for all 15 years was modelled using Spline function and a norm about 0.001. This construction under MATLAB has developed a lot of intermediary data which are accepted as being real and appropriate the continuous state function X_C, in the same time for S1 and S4 stations.

The next step was to consider the difference between the level of (DO) in S1 and for S4 in a data table in EXCEL. The last step was to import the data in software "Table Curve 2D", and to consider Chebyshev Std Polynomial Order 20 for these difference.

Analyzing the map, some critical periods have been identified, around 1993-1994 and one more important around middle of 1999 and the end of 2001, when S1>S4. These periods could be exactly identified and analyzed from the point of view of the impact of wastewater source from Braila and Galati region, (Figure 10).

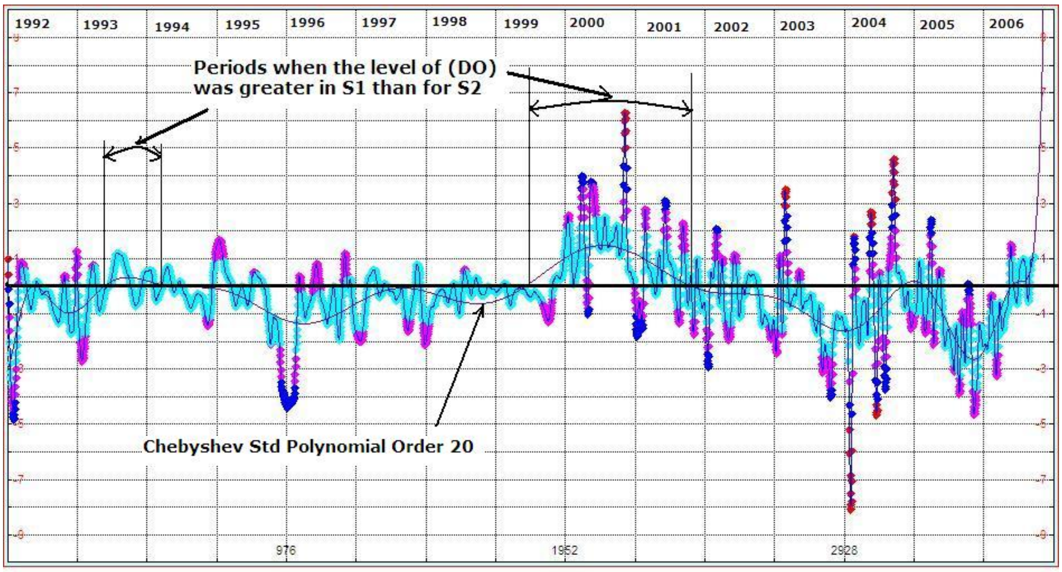


Fig. 10. The map of difference between the level of (DO) for S1 and the level of (DO) for S2, using values modelled in conform to the proposed mathematical analysis and processed with TC2D software.

4. RESULTS AND DISCUSSION

To achieve mathematical model indicators of the Danube River water quality were taken into account and thus achieved a conceptual specific model for time series analysis. Thus, in order to define the status function of the quality parameters, using the notion of time series, and framing their analysis in the observable space of parameters.

State functions must model the behaviour of quality parameters of the oxygen group, however, the model is intended to be a model applicable to all quality data, in order to be a flexible and generally applicable model for the methodological study of quality parameters and their evolution in time.

By analyzing a database containing rich data, one could make connections between the evolutions of the parameters and should identify the impact of the polluting sources in the segment studied, and the river could be characterized by its ability to drain some waste water and specific pollution sources. Analyzing the evolution of quality parameters one could overlap the maps and identify the period when something enters well in the station S1 and goes out wrong in the station S4.

The polynomial function maps could be used for this analysis. In this mode one could give the correct impact of significant urban agglomerations included in the segment of Danube River analyzed.

Some evolution could be seen well by using the polynomial forms of proper state function X_C for periods shorter than one year but others must be analyzed with other equation form, more complex than the polynomial form.

5. CONCLUSIONS

One of the most important conclusions of this study is to implement the methodology of analysis of the water quality data, taking into consideration not only one data in a day but accept that this data is the result of an evolution and in general is influenced by a short history of data. By including a measurement of one data in a string, one could include a data in an evolution, in a trend of this evolution. In the meantime, we could see better the anomalies of measurement, taking into consideration the polynomial form considered for one year, for one station of monitoring and assume that some data could not be very out of values of the polynomial function proposed. So, if this difference between the determinate data and the value of polynomial form is greater than 2 units and no other significant influences are observed, than the measurement must be repeated. Analyzing the database, usually the conclusions could be of the form “After 1991 the quality of the Danube River water has had a good trend”, [8]. But, the proposed methodology shows a lot of short period of time when this affirmation is not conform to this conclusion. There have been discovered many periods of time when for a period greater than 30 days (DO) it was not enough to maintain the aquatic life.

REFERENCES

[1] Muraru, C.V., MATLAB Ghid de studiu, Editura EduSoft, Bacau, 2006.
[2] Colojoară, A., Serii de timp, Editura Universităţii din Bucureşti, 2007.
[3] Popovici, C.I., Analiză numerică cu MATLAB, Casa de editură Venus, Iaşi, 2007.
[4] Trâmbiţaş, R.T., Analiză numerică-O introducere bazată pe MATLAB, Presa Universitară Clujeană, Cluj-Napoca, 2005.
[5] Păltineanu, G., Matei, P., Trandafir, R., Analiză numerică, ebook, http://civile.utcb.ro/cmat/cursrt/an.htm.
[6] Sârbu, I., Modelarea proceselor ecologice, Note de curs, Univ. Lucian Blaga Sibiu, 2009.
[7] Antohe, V., Stanciu, C., Modeling and simulation of quality indicators of surface waters, Environmental Engineering and Management Journal, Vol.8, No.6, 2009, p.1422.
[8] Report ICPDR 2005 WFD Roof Report 2004, ICPDR Doc IC/084 2005.