

Research of Data Sequences of Groundwater Levels with Gaps

Mariia Tymkiv^{1*}, Dmytro Kasiyanchuk¹

¹ Ivano-Frankivsk National Technical University of Oil and Gas, Ukraine

* Corresponding author's e-mail: maritymkiv@gmail.com

ABSTRACT

Investigating the regime of groundwater levels is one of the main stages of hydrogeological research, which allows quantitative characterization of the process of formation and change of hydrogeological conditions in time. To date, there has been a problem in filling the series of groundwater levels (GWLs) that have gaps. During the formation of time series of data, there are usually gaps in connection with the change in the technique of observing the GWL, the untimely filling of data rows, the lack of the ability to conduct such observations, and others like that. However, it is not possible to completely exclude rows with spaces, as this may lead to inaccurate data while developing forecast and constructing predictive maps of groundwater levels. Therefore, the task is to compare the existing methods of filling the data series with spaces in order to choose the optimum for further work, with their practical verification.

Keywords: groundwater, gaps, analysis

INTRODUCTION

There are many methods for filling the gaps, which occur in the time sequences. In their investigation, Little and Rubin [1990] outlined the modern methods and algorithms for processing the gaps in values in different tasks of statistical analysis. The paper of Dempster et al. [1977] widely presents the algorithm for calculating the maximum likelihood estimation in accordance with the incomplete data, which are presented due to the different degrees of data generation. Snytiuk [2006] made a vital contribution to the development of the statistical analyses in Ukraine. The methods of processing the data gaps were considered in the papers Bidiuk [2014] and Kuznietsova [2016], which offer classification for the methods, taking into account the varieties of the input data, types and formats of data, reasons of gaps caused by the influence of uncertainties of the surrounding world and the modeling object. We propose to apply the methods of intellectual analysis of data for processing the gaps in values and there is an example of filling the gaps in data with the methods of regression

analysis, by means of the forecast evaluations, in particular. Snytiuk [2006] describes the evolutionary method based on the composition using the neural network and the genetic algorithm. The technology of gaps renovation, which is offered by the scientific paper, does not require any restrictions connected with the linearity of model, parameter distribution, etc. Rossiev [1998] offered the method of sequential modelling of the data set with one-dimensional curve. The method is interpreted by the construction of a neuron conveyor for processing the data with gaps. The other possible interpretation shall be the iteration method for the main component parts and nonlinear factor analysis for data with gaps.

ASSIGNMENT OF TASK AND SELECTION OF TERRITORY

The purpose of the scientific paper is to analyze the methods of filling the time sequences, which have gaps, as well as creation of the new model to conduct the time forecast according to the groundwater levels for the area with the

different regime and type of groundwater. For this purpose, we analyzed the modern methods of filling data sequences; the methodology of hydrogeological data analysis with gaps is improved. We also offered the calculation of the reference sequences based on incomplete time sequences and creation of the new filled sequences. It is reflected in the regime of groundwater level.

The areas of Rivno region, the northwestern part of Ukraine, are mostly located between two large platform structures – the Ukrainian shield and Volyn-Podillia platform and just a small area on the northeast outskirts of Rivno region is situated within the Prypiat topographic trench. In the sense of hydrogeology, the Rivno region is located in the area of three artesian basins of the groundwater: Volyn-Podillia, Prypiat, and Ukrainian basins of the fracture water.

OVERVIEW OF MAIN APPROACHES TO RESEARCH OF GROUNDWATER LEVELS WITH GAPS

Scholars have suggested various methodologies for filling the time sequences with gaps. Nevertheless, nowadays, we do not have a single methodology for filling the data. Therefore, the main purpose of the thesis shall be a selection and justification of the methodology for filling the time sequences of data with gaps. One of the tasks for further work is filling the sequences with gaps in such manner that the changes introduced can be shown correctly. We may highlight the basic methods for the fast filling the data [Kuznietsova 2016].

- The methods of multiple filling stipulate filling a gap by several values. The essential drawback of the methods of one-time filling, according to their researchers, is that the ordinary formulas result in the systematic underestimates of dispersion, even if the missed values are derived by using the correct model. We obtain the correct estimates of dispersion during the multiple filling. Such estimates can be derived by the ordinary methods of analyzing the complete data.
- Filling in accordance with regression means that the data gaps provided by the regression of missed values for a particular variables object are filled by those data which are known.
- Filling without selection means that the gaps are filled by the constant data from the external source, e.g. by the data from the previous studies at the same place.

- The method of substitution is characterized by replacement of data at the stage of observation and at the collection of data exactly. The essence is in the substitution of the object having the missed required data by another object, which is not included in the selected information. However, such selected information should not be considered as complete and reliable, as the selected data also may systematically differ from the correct data.
- The comparison of methods is based on the several methods. While working with data one may need to compare one or another method based on a real observation. For example, we may combine the filling with selection and the filling by regression [Bidiuk 2014].
- The gaps are filled by the average value taken from the five values in the sequences, taking into account the average values in years and months. This method is the most well-known and common among the scholars. It is effective and shows the good accuracy of results during experiments. However, this method is the most effective at the single gaps.
- Filling in accordance with the regression consists in filling the data gaps, which are stipulated by regression of missed values for a particular variables object, by the known values. A major drawback of such method is that the forecast quality of renovation of the missed values depends on the choice of the regression model, which is used as a basis.
- Exclusion of sequences with gaps is the method, which is easy to apply. However, the necessary condition for its use is that the gaps in data should be completely random. It is important that this method should be usually used with the small number of gaps in the table. The main drawback of such approach is stipulated by the loss of information when excluding the incomplete data [Little and Rubin 1990, Dempster 1977].

METHODOLOGY

The choice of factors for analysis of groundwater level, as it is known, mainly depends on the meteorological conditions, geological and anthropogenic factors. Altovskiy [1954] offered the following classification of factors that influence over the groundwater regime (Table 1).

The main factors which form the regime of groundwater levels and should be taken into consideration during its study and forecast, are:

Table 1. Classification of factors of the groundwater regime

Genetic group of factors	Group of factors		
	Changed slowly	Changed quickly	Episodic
Climatic	Types of climate	Atmospheric precipitations, evaporations, atmospheric pressure, air temperature	Heavy rains, thaws
Hydrological	Types of regimes of the rivers, lakes and seas	Floods, surf, influx and reflux	Flood tides, Wind set up and set down, ice jams
Geological	Geological structures, tectonic movements, internal energy and cinematic of the Earth, radioactive elements of rocks		Earthquakes, volcano eruptions, mud volcanoes, earth faults
Soil-forming	Types of soil-forming process		
Biogenic	Perennial canopy transpiration	Seasonal canopy transpiration	
Man-made		Drying up, water supply	

climatic (meteorological) ones, atmospheric precipitations and the air temperature [Bilinska and Tymkiv 2016]. According to Konopliantsev, Semenov [1963]; Kovalevskiy, [1983]; Ruban, Shynkarevskiy [2005], the analysis of a previous research of long-term variability of groundwater levels shows that first of all, it is expedient to consider such main characteristics of metrological factors creating regimes: annual precipitation, precipitation in cold season (from November to February), average annual air temperatures, and average summer air temperatures. These parameters enable us to offer the quantitative characteristics for complex influence of factors, which determine the increment of groundwater in each particular year. The changes in meteorological factors, which directly influence the fluctuations of groundwater levels, are mostly stipulated by the cosmogenic factors and the changes in solar activity in particular. A great number of researchers [Konopliantsev, Semenov, 1974; Kovalevskiy, 1976, Pona et al. 2016] specified the regular relationship between the long-term fluctuation of groundwater levels and the changes in solar activity. The hydrogeological factors of the regime forming are distinguished as a separate group. This fact is connected with specific character of their influence [Davybida, Kuzmenko, 2011].

After selecting the factors that influence the regime of groundwater level, we conducted the analysis of the output data. An aggregate of the time sequences that are the data of observation according to groundwater levels at certain aggregate of the observation wells was considered.

The existing approaches to the analysis of the time sequences of groundwater levels require the

clearer structuring to provide the reliable informative nature of their forecast. Figure 1 shows the methodology for the analysis of the hydrogeological data with gaps for the certain period [Tymkiv et al. 2018].

The choice of area is the first and the most important stage. As the hydrogeological structure of Ukraine is quite complicated, it is expedient to carry out the research within the river basins, as the hydrogeological basins by their territorial components are commensurate with the hydrological structure.

The next stage is the creation of a database, which will contain the general data and hydrogeological characteristics pertaining to the investigated well. Since the present state of wells, where we can observe that the groundwater level is satisfactory, the creation of the database for the existing wells is one of the priority steps to establish a new network for observation over the groundwater as a whole.

First, we should pay our attention to the importance of the certain gaps that were calculated due to the great number of statistical methods for filling the time sequences. Perhaps, this problem is the greatest one, as the scholars do not take into account the correlation and statistical comparative characteristics or recognize them in comparison with the average value of the initial unfilled sequence, rather than the data sequences taken before and after filling [Kasiyanchuk et al. 2016].

The following paragraphs of the proposed methodology for analyzing the hydrogeological data with gaps lay the foundation for the time forecast of the groundwater level. At the same time, the new approaches to filling the data with gaps are considered.

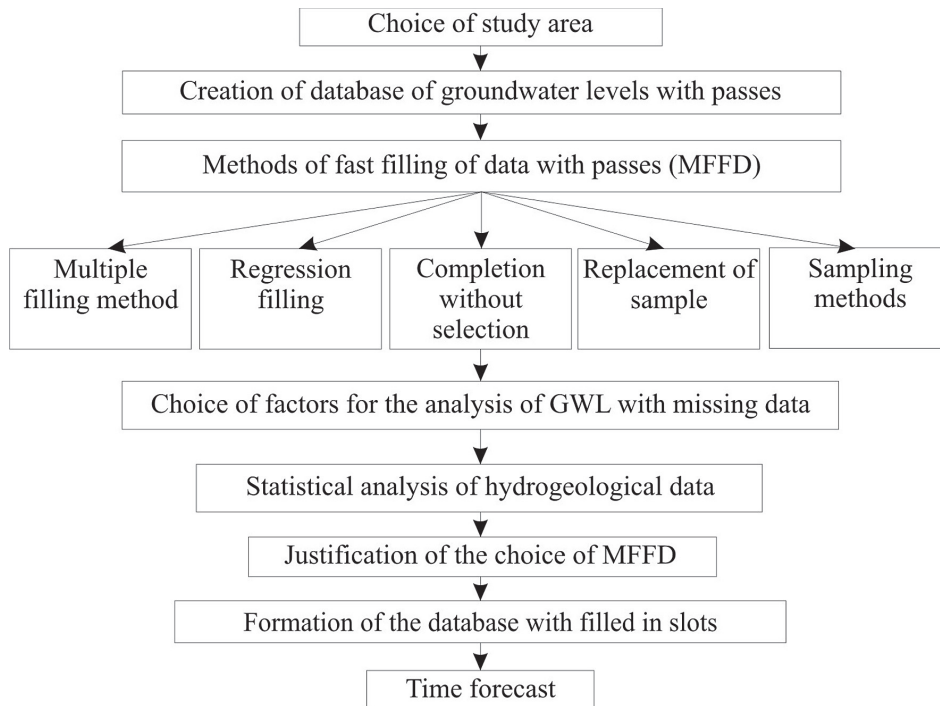


Figure 1. Methodology of analysis of hydrogeological data with gaps

CHOICE OF DATA FOR ANALYSIS

The analysis of the data with gaps was performed on the example of well No 255310011 of the confined aquifer in the area with the undisturbed regime in the village of Zirne and well No 255030066 of the confined aquifer located in the area with the disturbed regime, in the urban

village of Hoshcha, Rivne region. The main reason for the disturbed regime of groundwater is farming in this area. The well is located not far from the Sluch River, which is the tributary of one of the main rivers in the region – the Horyn River.

In order to conduct a complete analysis of the wells, we must have the qualitative data sequences of the groundwater levels. The data sequences

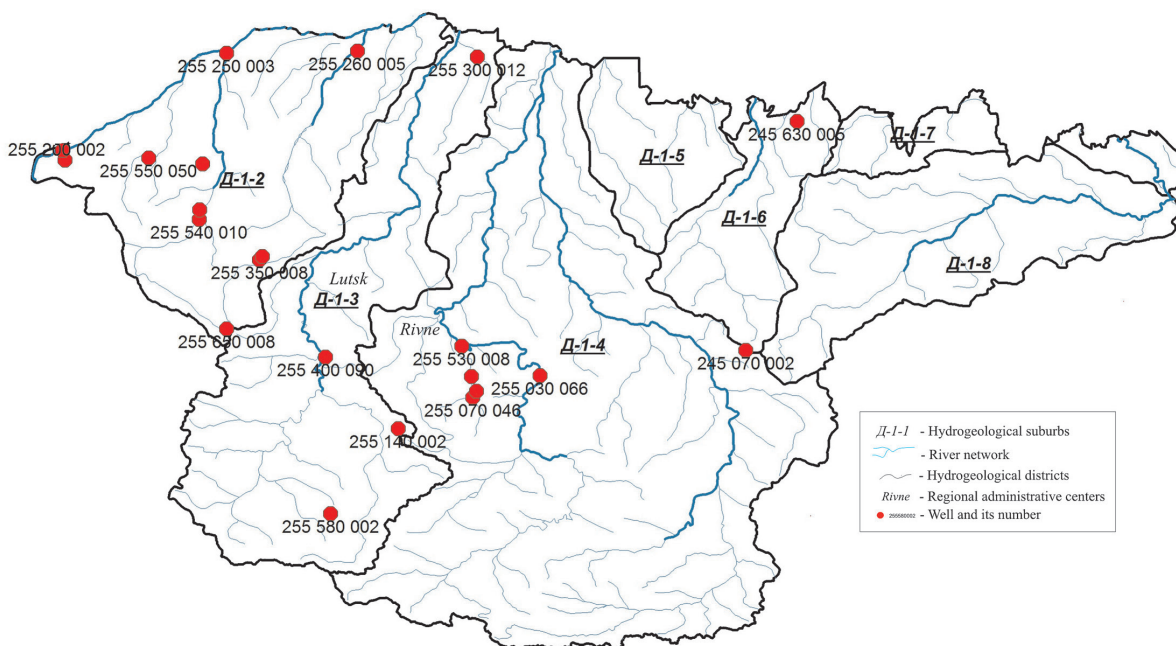


Figure 2. Map of wells on the territory of the Pripjat hydrogeological basin

should be statistically homogenous, without gaps, and with long-term period of observations, etc. Unfortunately, as we can see from the researched wells (Table 2) not all of these values are adhered to. There may be several reasons for this fact: non-systematic observations, absence of observations in general, error in entering data, etc. Accordingly, all the outgoing data sequences of observations on groundwater levels have gaps (10.4% and 10.6% of gaps). Thus, there is the next step that is our personal filling of these gaps for further forecast.

RESULTS

In order to ensure the unity of displaying measurements, we should have a clear identity of units in calibration of all technical devices intended for measurement of the same physical value. This is achieved by means of precisely reproducing and preserving of units of physical values accepted at the International Conference on Measure and Weight and reflection of their sizes by means of measurements.

Special references are developed to reproduce the units under special conditions, where the direct reflection of the unit size according to the reference is technically impossible with the given accuracy (high pressure, temperature, frequency, etc.).

In their work, Ruban et al. [2000] presented the reference sequences for a complete time sequence (95%) taking into account the type of regime and groundwater depth. The novelty of this work was the construction of the reference sequence for such kind of sequence that is containing the gaps. The present reference sequence may be applied for the data sequence with any number of gaps.

Algorithm for renovation of gaps

The data analysis was conducted for each well individually within a single hydrogeological basin.

The first stage is to find the average monthly value and the standard deviation for the entire period of observations that were conducted on the well.

The equation of curve describing the annual (long-term) average dynamics of groundwater

Table 2. Outgoing data of observations on groundwater levels

a)

Well No	year/month	1	2	3	4	5	6	7	8	9	10	11	12
255030066	1996	3.86	3.65	3.07	2.23	1.79	2.18	2.74	3.07	3.21	3.23	2.97	3.31
	1997	3.45	3.43	3.80	3.92	3.53	3.53	3.51	2.92	2.68	2.67	-	-
	1998	-	-	2.99	3.26	3.45	3.85	2.65	2.03	1.99	1.58	1.59	1.21
	1999	2.04	-	0.44	0.63	0.96	0.82	1.22	1.48	1.64	1.68	2.02	1.78
	2000	1.29	0.67	0.14	0.27	0.78	0.33	0.75	0.55	0.16	0.80	1.56	1.30
	2001	0.67	0.72	0.43	0.60	1.27	1.56	0.60	0.28	0.85	0.86	0.76	1.14
	2002	1.27	0.63	0.60	0.80	1.04	0.97	0.75	1.84	2.03	1.98	-	-
	2003	-	-	1.97	2.11	1.38	-	1.02	1.14	1.38	1.66	1.75	2.08
	2004	-	-	1.02	1.43	1.50	1.93	2.09	1.41	1.65	1.56	1.56	1.92
	2005	1.84	1.97	2.12	0.84	0.63	0.85	1.18	1.08	1.19	1.30	1.44	1.16
	2007	2.53	2.17	1.58	1.29	1.58	1.77	2.41	2.24	1.79	1.56	2.22	2.30
	2008	-	1.90	1.89	1.66	1.03	0.89	0.74	0.70	0.75	0.92	1.20	1.41

b)

Well No	year/month	1	2	3	4	5	6	7	8	9	10	11	12
255310011	1996	1.87	1.87	1.84	1.27	1.59	1.75	1.85	1.86	1.87	1.88	1.86	1.69
	1997	1.68	1.64	1.67	1.61	1.64	1.61	1.62	1.59	1.67	1.62	-	-
	1998	-	-	1.16	1.11	1.41	1.56	1.33	1.35	1.29	1.16	1.14	1.21
	1999	1.02	1.18	0.92	1.17	1.4	1.65	1.64	1.46	1.47	1.48	1.43	1.22
	2000	1.32	0.99	1.11	1.07	1.44	1.63	1.61	1.51	1.63	1.67	1.68	1.6
	2001	1.16	1.11	1.21	1.48	1.48	1.55	1.63	1.57	-	1.56	1.59	1.63
	2002	1.62	1.17	1.17	1.35	1.56	1.65	1.79	1.9	2.05	1.98	-	-
	2003	-	-	1.08	1.36	1.65	1.75	1.78	1.75	1.87	1.87	1.79	1.64
	2004	-	-	1.02	1.21	1.4	1.6	1.72	1.86	1.85	1.87	1.88	1.38
	2005	1.49	1.39	1.14	0.95	1.03	1.32	1.51	1.64	1.67	1.75	-	-
	2007	1.62	1.25	1.11	1.31	1.51	1.63	1.39	1.23	1.32	1.57	1.6	1.74
	2008	1.61	1.23	1.15	1.01	-	1.4	1.5	-	1.58	0.98	1.13	1.38

level has the form of equation taking into account the seasonal variations of groundwater levels (1):

$$f(x) = a_0x^4 + a_1x^3 + a_2x^2 + a_3x + a_4 \quad (1)$$

Mean-square (standard) deviation will look as follows:

$$\delta_{N_i}^m = \sqrt{\frac{(X_{N_i}^m - \bar{X}_{i_N}^m)^2}{n}} \quad (2)$$

where: X – groundwater level
 m – month of observation of groundwater level
 N – number of years of groundwater level observations
 i – value of groundwater in N-year
 n – number of groundwater values

Calculation of curve function describing groundwater level (reference sequences) was carried out using the following formula:

$$f(x(\delta_y^m)) = a_0(\delta_y^m)^4 + a_1(\delta_y^m)^3 + a_2(\delta_y^m)^2 + a_3(\delta_y^m)^1 + a_4 \quad (3)$$

$$\varepsilon = \frac{f(x(\delta_y^m)) + \bar{X}_m^y}{2} \quad (4)$$

Ruban et al. [2000] offer a reference sequence, which is structured in accordance with the hydrogeological deep gradation of groundwater level, as a method for filling the time sequence. The reference sequence is given as an average value of multi-year values of levels without gaps. In this case, the author takes into account the dynamics of water levels in the well within the months.

Unfortunately, the wells 255310011, 255030066 do not have the full sequence of groundwater level, and for this reason, the above-mentioned method for the calculation of the reference sequence is unacceptable.

The problem of interpretation of the data with gaps is very important while choosing the statistical methods for the calculation of absent numerical characteristics of incomplete time sequence. If we consider the numerical changing in dynamics of any sequence while simulating a different volume of gaps, an issue that the existence of the reference sequence as such can be replaced by the method of filling by five nearest average values, arises. Nevertheless, in such event, the time presentable appearance of the sequence is lost due to the impossibility of observance over long-term fluctuations.

The search of solution for the task, which is connected with use of known mathematic methods of filling, becomes a new problem because of the threat of absence of presentable appearance of the sequence having a multi-year monthly component.

The existing methods generalize the gaps as the aggregate of any values, years, or months. This, in turn, leads to the false representation of a sequence filled by one of the methods.

At the same time, the correlative characteristic of sequences is determinative, and from the point of view of how the average value can present the time change of groundwater level. For this purpose, the reference sequence characterizes the change, the dynamics of groundwater level. It can be traced through the time forecast.

The gradation of the multi-year groundwater levels, taking into account over 10% of the volume of gaps from the total number of data, makes us search for the new approaches to study the method for the calculation of the reference sequence (ε).

According to statistics, the nature of changing values in estimation of their reliability reflects a parameter of mean-square deviation (σ). It was chosen because the ordinary sequence of average values cannot reflect its physical significance due to the present gaps. The approaches using statistical methods for the calculation of the reference sequence (ε) are very important from the point of view of the presentable appearance of new (reference) values with the initial sequence. Formula 2 presents the calculation of mean square deviation of the well. Accordingly to the values of the new sequence (mean square deviation), we can calculate the values at the point of observation as a function of curve of the output averaged sequence (formula 3).

The reference sequence (ε) is a sequence, which reflects the average multi-year dynamics of change in any value and deviation from its average value. Formula 4 represents a calculation of the reference sequence (ε) of groundwater level as a component of the afore-mentioned values.

Figures 3–4 show the dependency diagram of the standard deviation, the average and reference value for the time sequence with gaps and for the same sequence filled with the reference value (10%, 30%, 50% of gaps).

For better representation of the results, Figures 3–4 should be presented in the form of table with correlation coefficient of the relationship value.

Reviewing the results of correlation analysis, it is worth noticing that the approaches to calculation of values of gaps in the sequences of groundwater levels are relevant. There is a high correlation relationship at any value of the data confidence (significance) level. However, we can trace a clear tendency, which explains the correctness of the calculation method for the reference sequence having the data with gaps. As we can see from the tables, the more gaps we have, the lesser the correlation value for the method of five nearest values is. The correlation values remains consistently high for the reference sequences.

Tables 5–6 present the results of the correlation analysis pertaining to the relationship between the renovated data, reference sequence, and the average value of the filed sequence according to well No 255310011. The reference sequences were constructed on the basis of annual average values and constructed line of the trend (Fig. 2, 3). The value of determination coefficient shows that the linear regression equation complies with the selected data.

The procedure for analyzing the sequences of groundwater level using the reference sequence is based on the calculation of mean square deviations in the well for each month. A polynomial line characterizing the average annual values of groundwater level is constructed. The line is normalized to the desired values according to the relevant form, where the undetermined values are

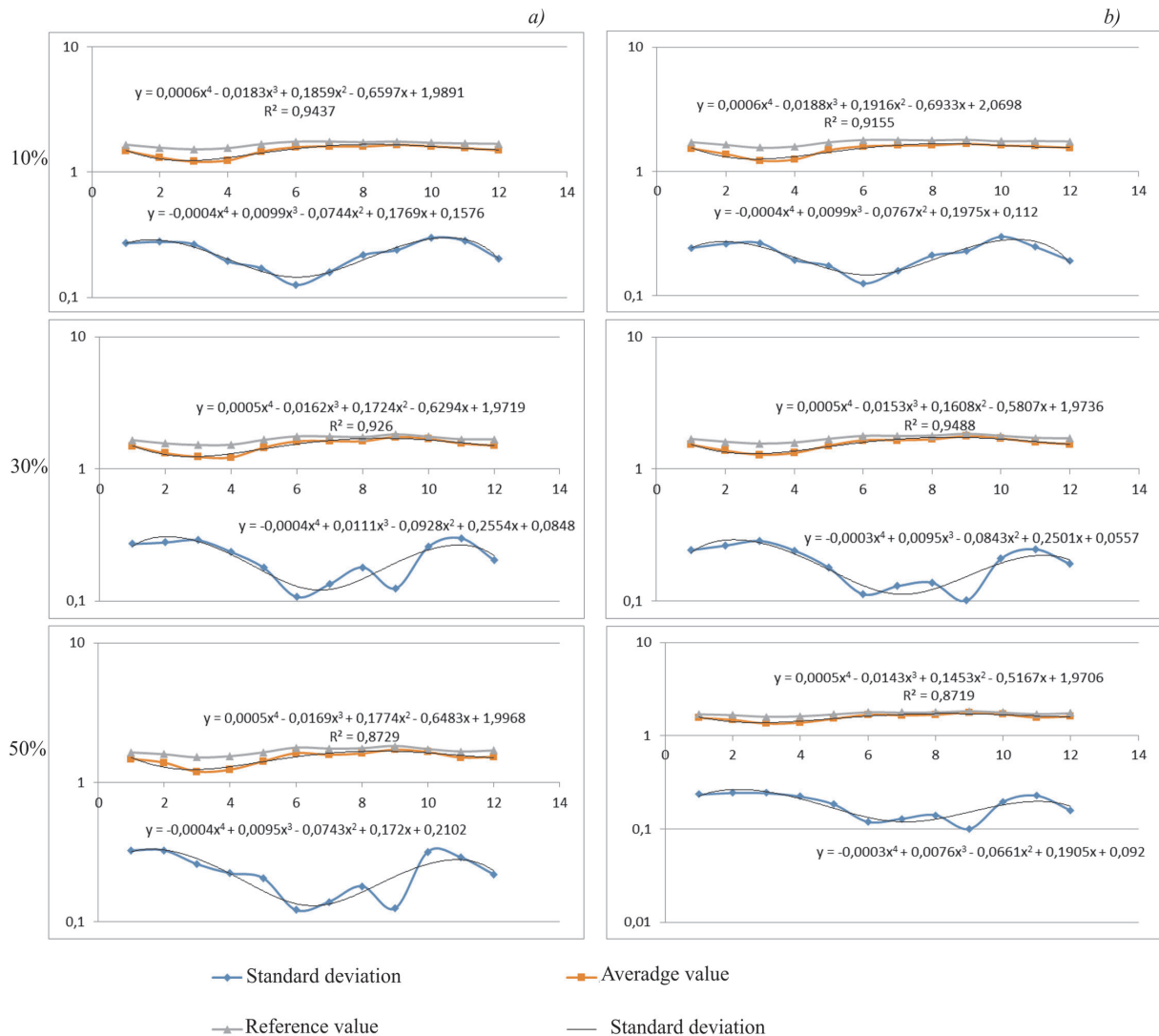


Figure 3. Dependency diagrams of the standard deviation, average, and reference value for well No 255310011: a) time-sequence with gaps; b) time sequence filled with the reference value

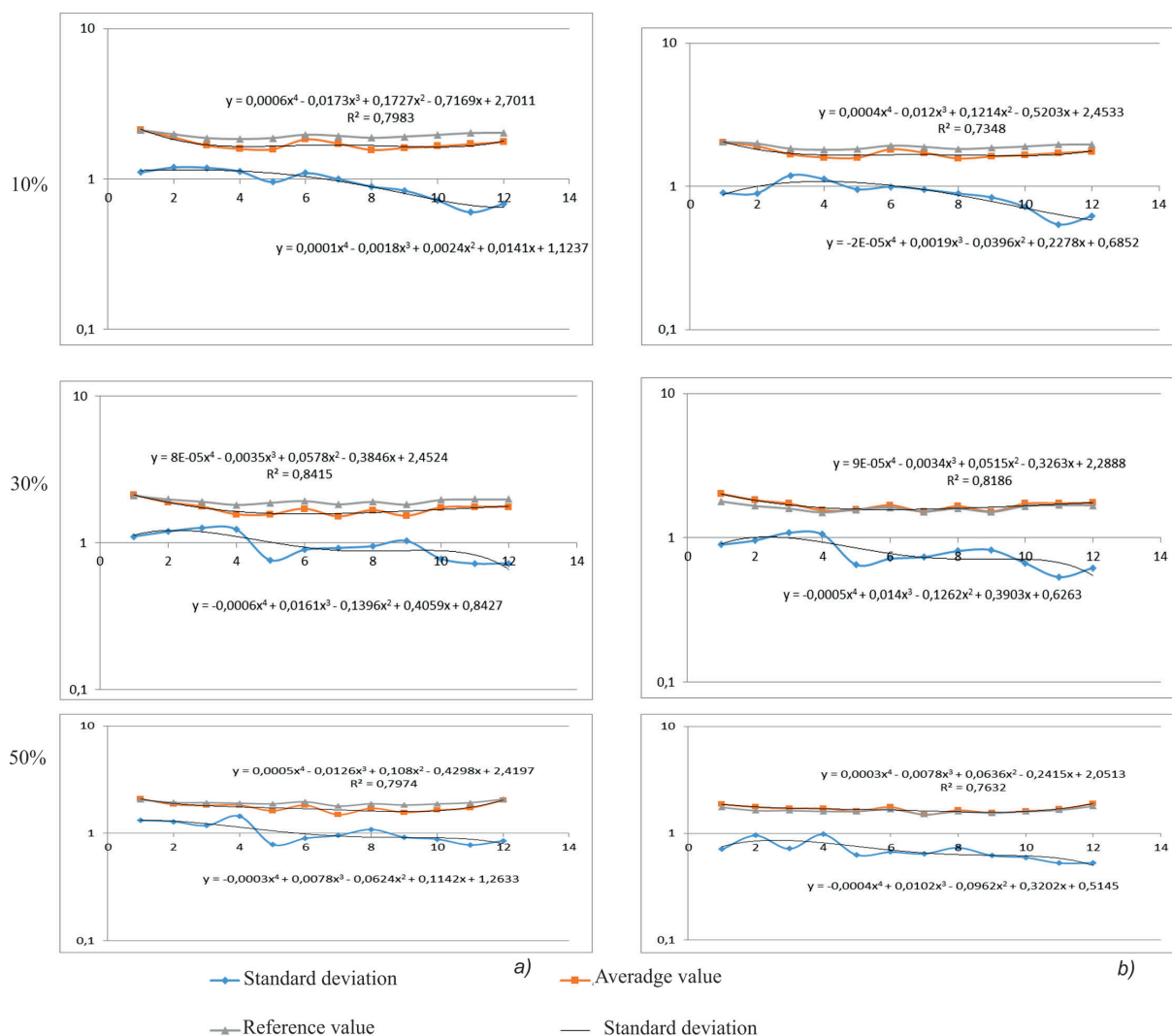


Figure 4. Dependency diagrams of the standard deviation, average, and reference value for well № 25503066: a) time-sequence with gaps; b) time sequence filled with the reference value

Table 3. Correlation coefficient of the relationship value

WellNo 255310011 Undisturbed regime (natural)	Average value of the output sequence / Average value of sequence filled with reference value	Averagevalue (methodof fivenearest values)/ Average value of the output sequence	Averagevalue (methodof fivenearestvalues)/ Average value of sequence filled with reference value
10%	0.989698	0.994432	0.9925
30%	0.995704	0.993178	0.993488
50%	0.996030121	0.973488222	0.975916126

Table 4. Correlation coefficient of the relationship value

Well No 25503066 Disturbed regime (anthropogenic)	Average value of the output sequence / Average value of sequence filled with reference value	Averagevalue (methodof fivenearest values)/ Average value of the output sequence	Averagevalue (methodof five nearestvalues)/ Average value of sequence filled with reference value
10%	0.990755	0.962449	0.947818
30%	0.99729834	0.90455801	0.8926588
50%	0.982273	0.816875	0.853012

Table 5. Value of correlation between the reference sequence, the averaged values and according to years for well No 255310011 (a) and filled time sequence (b)

a)

Correlation coefficient	Reference sequence	Average values	Values according to years
1996	0.367228	0.506809	0.442679
1997	-0.18812	-0.22052	0.087654
1998	0.411471	0.253733	0.249687
1999	0.854882	0.777295	0.053757
2000	0.909824	0.916996	0.206069
2001	0.739865	0.709323	0.10233
2002	0.896322	0.924124	0.265913
2003	0.949759	0.948178	0.199822
2004	0.857971	0.926108	0.225159
2005	0.622477	0.795603	0.393354
2007	0.46527	0.467321	0.132994
2008	0.354794	0.42245	1

b)

Year/month	1	2	3	4	5	6	7	8	9	10	11	12
1996	1.87	1.87	1.84	1.27	1.59	1.75	1.85	1.86	1.87	1.88	1.86	1.69
1997	1.68	1.64	1.67	1.61	1.64	1.61	1.62	1.59	1.67	1.62	1.46	1.43
1998	1.31	1.39	1.16	1.11	1.41	1.56	1.33	1.35	1.29	1.16	1.14	1.21
1999	1.02	1.18	0.92	1.17	1.4	1.65	1.64	1.46	1.47	1.48	1.43	1.22
2000	1.32	0.99	1.11	1.07	1.44	1.63	1.61	1.51	1.63	1.67	1.68	1.6
2001	1.16	1.11	1.21	1.48	1.48	1.55	1.63	1.57	1.83	1.56	1.59	1.63
2002	1.62	1.17	1.17	1.35	1.56	1.65	1.79	1.9	2.05	1.98	1.69	1.63
2003	1.58	1.24	1.08	1.36	1.65	1.75	1.78	1.75	1.87	1.87	1.79	1.64
2004	1.53	1.31	1.02	1.21	1.4	1.6	1.72	1.86	1.85	1.87	1.88	1.38
2005	1.49	1.39	1.14	0.95	1.03	1.32	1.51	1.64	1.67	1.75	1.78	1.49
2007	1.62	1.25	1.11	1.31	1.51	1.63	1.39	1.23	1.32	1.57	1.6	1.74
2008	1.61	1.23	1.15	1.01	0.78	1.4	1.5	1.18	1.58	0.98	1.13	1.38

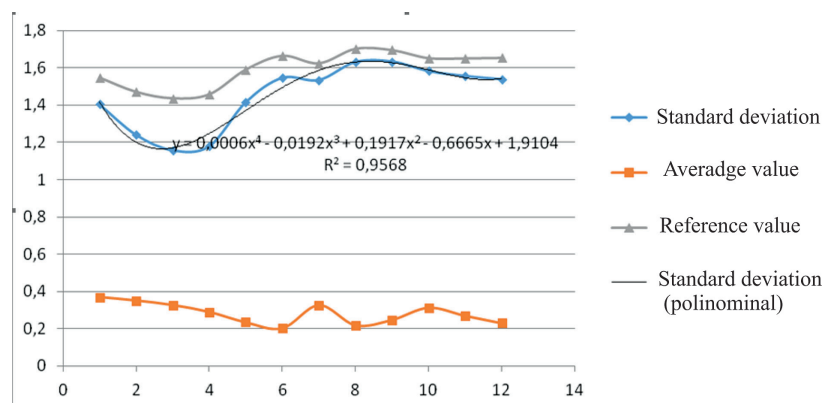


Figure 5. Dependency diagram of annual average and reference sequence of groundwater level for well No 255310011

equal to the relevant value of the mean square error of a month. Forming of the referent sequence permits us to describe clearly and trace the change in dynamics of groundwater level for each month. Such approach enables to conduct the constructive analysis of sequences with gaps and normalizing the data on the wells to the reference ones.

As it can be seen from Table 5, the correlation values are non-unique for certain sequences. In particular, the correlation in 1997 indicates that the measurements of groundwater levels do not correspond to the general dynamics as a whole. While analyzing the correlation values between the reference and average ones

Table 6. Value of correlation between the reference sequence, average values and according to years for well No 25503066(a) and filled time sequence (b)

a)

Correlation coefficient	Reference sequence	Average values	Values according to years
1996	0.587897	0.640056	0.174968
1997	-0.36607	-0.09567	0.277534
1998	-0.43915	-0.16382	0.026908
1999	0.649263	0.495605	-0.41751
2000	0.6399	0.575809	-0.07068
2001	-0.18798	0.039627	-0.18556
2002	0.064386	-0.07302	-0.61732
2003	-0.43313	0.069924	0.416964
2004	0.422862	0.150164	-0.5864
2005	0.326881	0.557188	0.543241
2007	1	0.578626	-0.22574
2008	-0.22574	0.189174	1

b)

Year/month	1	2	3	4	5	6	7	8	9	10	11	12
1996	3.86	3.65	3.07	2.23	1.79	2.18	2.74	3.07	3.21	3.23	2.97	3.31
1997	3.45	3.43	3.8	3.92	3.53	3.53	3.51	2.92	2.68	2.67	2.17	2
1998	2.65	1.99	2.99	3.26	3.45	3.85	2.65	2.03	1.99	1.58	1.59	1.21
1999	2.04	1.15	0.44	0.63	0.96	0.82	1.22	1.48	1.64	1.68	2.02	1.78
2000	1.29	0.67	0.14	0.27	0.78	0.33	0.75	0.55	0.16	0.8	1.56	1.3
2001	0.67	0.72	0.43	0.6	1.27	1.56	0.6	0.28	0.85	0.86	0.76	1.14
2002	1.27	0.63	0.6	0.8	1.04	0.97	0.75	1.84	2.03	1.98	1.15	1.54
2003	1.44	0.92	1.97	2.11	1.38	1.37	1.02	1.14	1.38	1.66	1.75	2.08
2004	1.63	1.35	1.02	1.43	1.5	1.93	2.09	1.41	1.65	1.56	1.56	1.92
2005	1.84	1.97	2.12	0.84	0.63	0.85	1.18	1.08	1.19	1.3	1.44	1.16
2007	2.53	2.17	1.58	1.29	1.58	1.77	2.41	2.24	1.79	1.56	2.22	2.3
2008	1.11	1.9	1.89	1.66	1.03	0.89	0.74	0.7	0.75	0.92	1.2	1.41

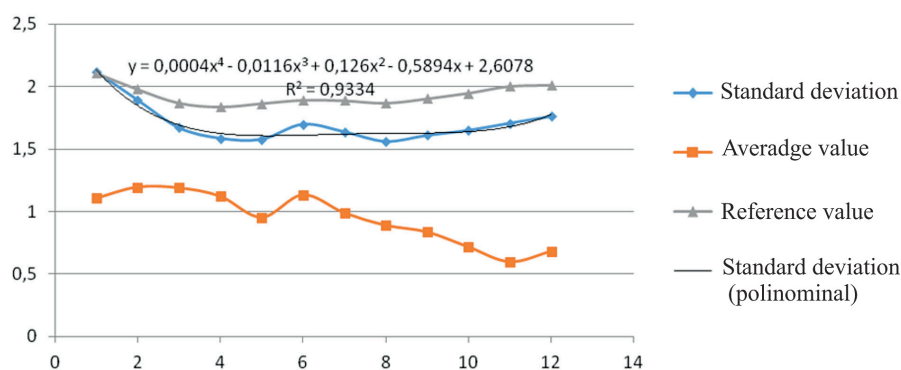


Figure 6. Dependency diagram of annual average and reference sequence of groundwater level for well No 255030066

(which are filled by methods for analyzing data with gaps), we may say that the first sequence is the most important. The significant fluctuation of groundwater levels within the wells for the selected period is reflected by slight correlation relationships according to years. This is mainly due to the influence of the main factor, the

amount of precipitation. It is pointless to trace the dynamics of multi-year change of groundwater levels for the certain well because of limited data quantity. For that purpose, we should consider other factors, which have a longer time sequence and may characterize the multi-year changes of groundwater level.

CONCLUSIONS

1. The observation network for monitoring groundwater levels with gaps does not allow a thorough approach to the forecast of their values in time. The gaps, which are present in the sequences of annual observations, require to be filled on the ground of statistical analysis of the well data.
2. As the results show, the selection of methods of filling data with gaps is many-valued. In particular, there is a statistical correlation dependency for the comparison between the multiple data filling and the reference sequence.
3. Researching the hydrogeological sequences with the methods of regressions and filling the selected information by the average value do not have a significant reliability. The best method is to conduct analysis of the selected information for filling the gaps by the average value among five ones and by the value from the reference sequence for the selected information concerning a certain well.
4. Taking into account the variability of dynamics of change in groundwater level, the calculation of the reference sequence for the well is the most optimal decision for conducting analysis of the data sequences. Determination of a sequence of long-term observations by presenting the dynamics in the form of a single – reference sequence allows representing the time change in physical values, as a possible forecast model of fluctuation of any calculated value.
5. The dynamism of groundwater level and the multifactor data, which need to be taken into consideration for the time analysis, should be considered in future due to using the other multi-year time sequence (solar activity, etc.).

REFERENCES

1. Bidyuk P.I., Korshevnyuk L.O. 2014. Models and methods of applied statistics. National Technical University of Ukraine “Kyiv Polytechnic”, pp. 722.
2. Bilinska, Y., Tymkiv, M. 2016. GIS modeling of the groundwater levels to the intensive flooded areas (on the example of the Zhytomyr Region). *Geoinformatics 2016 – XVth International Conference on Geoinformatics – Theoretical and Applied Aspects*.
3. Davybida L.I., Kuzmenko E.D. 2011. Regularities of subsurface waters’ levels dynamics and factors of its forming in the dnipropetrovsk regions’ territory. *Geodynamics*, 1(10), 83–93.
4. Dempster A.P., Laird N.M., Rubin D. B.. 1977. Like lihood from Incomplete Datavia the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*. No. 1, 1–38.
5. Kasianchuk, D., Kuzmenko E., Chepurna, T. and Chepurnyi, I. 2016. Calculation of that environmental and geological landslide risk estimate. *Eastern-European Journal of Enterprise Technologies*. Vol. 1, No. 10(79), 18–25.
6. Kuznietsova N. 2016. Identification and processing of uncertainties in the form of incomplete data by methods of intellectual analysis. *System Research & Information Technologies*, No. 2, 104–115.
7. Kovalevsky V.S. 1983. Long-term variability of groundwater resources. Moscow, Nauka, pp. 205.
8. Konoplyantsev A.A., Semenov S.M. 1963. The natural regime of groundwaters and its regularities. *Gosgeoltekhizdat*, pp. 231.
9. Little J., Rubin L. 1990. Statistical analysis of data with spaces. Moscow: Finances and Statistics, 336 p. III . (Mathematical-statistical methods abroad).
10. Pona, O., Shtogryn, L., Kasianchuk, D. 2016. The analysis of the relationship between the phases of the Moon and the occurrence of landslides. *Geoinformatics 2016 – XVth International Conference on Geoinformatics – Theoretical and Applied Aspects*.
11. Ruban S.A., Nikolishina A.V, Zhukova V.I, Mukhtarova M.B. et al. 2000. Information support for hydrogeological estimations and forecasts of the regime of groundwater of Ukraine. (Book 1). Dnipropetrovsk, 29–36
12. Ruban S.A., Shinkarevsky M.A. 2005. Hydrogeological estimates and forecasts of the underground water regime in Ukraine: Monograph. Scientific proceedings of UkrSGRI, pp. 572.
13. Rossiev A.A., Gorbaniya A.N. 1998. Modeling data with the help of screens for restoring data flows. In the book “Methods of neuroinformatics-mats”, KSTU: Krasnoyarsk, 6–22
14. Snityuk V.E. 2006. Evolutionary method for data retrieval. *Collected Works of the VI International Conference “Intellectual Analysis of Information”*, 262–271.
15. Tymkiv M., Kasiyanchuk D., Danyliak V., Levitska M. 2018. Study of the series of data of groundwater levels with passes. *Geoinformatics 2018 – XVIIth International Conference on Geoinformatics – Theoretical and Applied Aspects*.