5th International Congress of Serbian Society of Mechanics
Arandjelovac, Serbia, June 15-17, 2015

NONLINEAR TIME SERIES ANALYSIS OF FLUID DYNAMICS: STOCHASTIC GROUNDWATER LEVEL OSCILLATION

Srđan Kostić[1], Iva Guranov[2], Nebojša Vasović[3]

[1]Department of Scientific Research and Informatics, Institute for Development of Water Resources "Jaroslav Černi", Jaroslava Černog 80, 11226 Pinosava, Belgrade, Serbia
e-mail: srdjan.kostic@jcerni.co.rs
[2]Chair of Fluid Mechanics, Faculty of Mechanical Engineering
University of Belgrade, Kraljice Marije 16, 11000 Belgrade, Serbia
e-mail: iva.guranov@gmail.com
[3]Department of Applied Mathematics, Faculty of Mining and Geology
University of Belgrade, Đušina 7, 11000 Belgrade, Serbia
e-mail: nebojsa.vasovic@rgf.bg.ac.rs

Abstract:
In present paper we analyze dynamics of groundwater level oscillation recorded at two piezometric stations in Pancevacki rit: ''Cuvarnica'', and ''Borca'' in period 2007-2013. Primary goal of the performed research was to determine the character of the main mechanism behind these oscillations, which could further serve as a solid base for creating an appropriate prediction model. Dynamics of the recorded time series is examined using methods of nonlinear time series analysis and delay embedding theorem. After embedding the observed time series into three-dimensional phase space with embedding delay τ=4, results of surrogate data testing showed that analyzed time series originate from a stationary Gaussian linear process that could be distorted by a monotonic, instantaneous, time-independent nonlinear function. This is further confirmed by low values of determinism coefficient and corresponding vector field composed of vectors of different length, indicating high level of stochasticity in the observed data. The obtained results provide solid foundation for future research, with the final goal of creating an appropriate prediction model.

Key words: groundwater level, embedding theorem, surrogate testing, determinism, stochasticity

1. Introduction

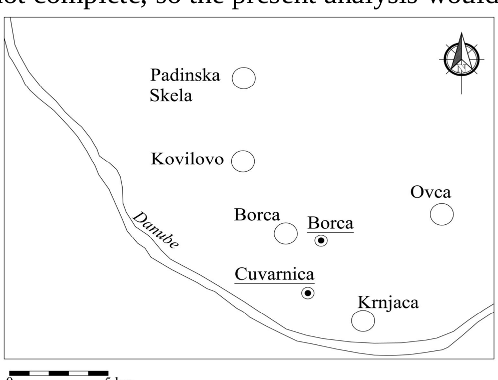
Groundwater dynamics is typically considered as one of the main parameters that control the slope stability. In particular, sudden increase of groundwater level due to the fast infiltration from the heavy precipitation or snow melt decreases the effective values of normal stress in soil and increases the hydrodynamic forces, which could further trigger landslides or similar colluvial processes. Common measures of remediation which are taken against such abrupt changes in groundwater dynamics include small concrete channels at the surface, which serve as collectors that enable precipitation flow outside of the endangered area. However, such measures are taken only for those terrains which are either already endangered by landsliding, or which are within the scope of engineering design, for which the soil stability needs to be ensured. In all other cases, which include slopes along the river valleys, artificial lakes, or within the urban areas, as it is the case in Belgrade, negative effect of groundwater level dynamics is often not propely taken into account. For such areas, groundwater level oscillation should be thoroughly studied in order to suggest and take appropriate measures, which would be both economic and provide safety against possible negative effects. For this purpose, certain prediction model should be developed, which could serve both as a temporal model (enabling short-or long term forecasting) and a spatial model, enabling the selection of the most endangered locations.

There were several attempts of predicting the groundwater level oscillation, mainly by applying neural network approach [1-6]. Although these models give reliable results, there is still no strong evidence about the character of the groundwater level oscillation, i.e. is it governed by purely deterministic process or there is still a certain stochastic component that mainly determines its behavior? We consider that this issue represents the first and the most important step in creating any reliable prediction model, since different mathematical approaches are used when dealing with deterministic or stochastic systems. For this purpose, it is necessary to analyze large number of recorded time series, which is rather convenient for territory of Serbia, since State Hydrometerological Survey [7] permanently tracks the groundwater level dynamics at 399 different stations – piezometers, where groundwater level oscillation is monitored since 1991. Within this research, we analyze recordings from two piezometers installed at Pancevacki rit (‘’Cuvarnica’’ and ‘’Borca’’).

In present paper, groundwater level dynamics is analyzed by using methods of nonlinear time series analysis. These methods were not previously used in the field of fluid dynamics, even though they were successfully applied for confirming the deterministic chaotic behavior of a simple periodically driven resistor-inductor diode [8], or in some other fields of geophysics, like seismology [9] or blast-induced vibrations [10]. These past studies have proved that nonlinear time series analysis methods have vast potential in studying the various types of experimentally recorded time series.

2. Experimental data

Time series which are analyzed in present paper are recorded at two piezometers at Cuvarnica and Borca station in Pancevacki rit, for the period 2007-2013 (Figure 1). These recordings were chosen as a preliminary case study, since method of nonlinear time series analysis is for the first time applied in fluid dynamics. Such short observation period is chosen since older recordings are not complete, so the present analysis would not be able to provide reliable results.



Station	Borca	Cuvarnica
Latitude	44^{0}52’33’’	44^{0}51’16’’
Longitude	20^{0}28’25’’	20^{0}27’49’’
Absolute height (m)	72.52	71.88
Depth of the installed piezometer (m)	12.00	12.00
Starting date of observation	01/01/1994	01/01/1994

Fig. 1. Distribution and main technical properties of the observed piezometric stations in Pancevacki rit.

Observation of groundwater level was performed every 5 days at Cuvarnica station and every 10 days at Borca station. Recorded time series of groundwater level oscillation are given in Figure 2. For days without observation, we used medium value between the succeeding and preceding recordings. Here we consider only the observations in the period 2007-2013, since there are no observations at these stations in 2005, and for first three and six months in 2006 at Borca and Cuvarnica station, respectively.

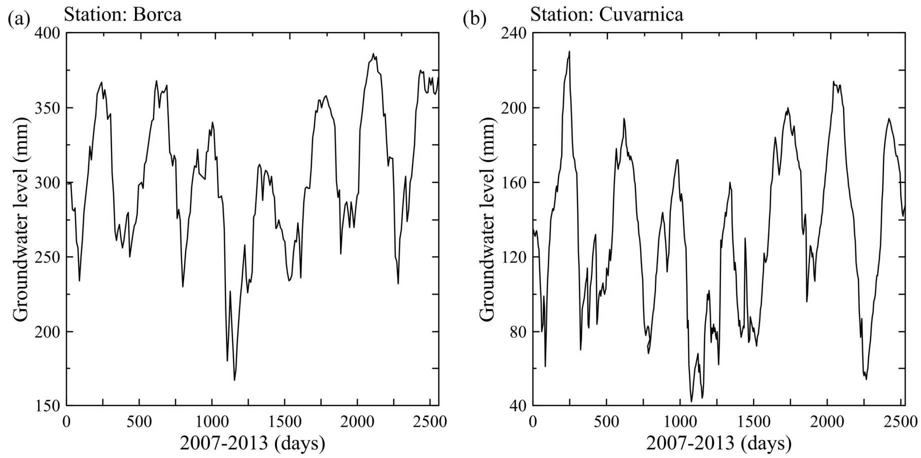


Fig. 2. Recorded time series of groundwater level oscillations at the following stations in Pancevacki rit: (a) Borca, (b) Cuvarnica. It is clear that both of the observed time series have similar trend.

From the hydrogeological viewpoint, investigated area belongs to ‘’Pancevacki rit’’ hydrogeological province, which represents alluvial plain of Danube and Tamis. The observed unconfined (phreatic) aquifer occurs within the layer of Quaternary clayey sands, which represent hydrogeological collector-reservoir of integranular porosity. Therefore, groundwater is commonly replenished by precipitation, which directly controls the dynamics of level oscillation [11].

3. Surrogate data analysis

Surrogate data analysis is performed by assuming that the observed data belong to some class of stochastic systems, i.e.: (1) data are independent random numbers drawn from some fixed but unknown distribution; (2) data originate from a stationary linear stochastic process with Gaussian inputs and (3) data originate from a stationary Gaussian linear process that has been distorted by a monotonic, instantaneous, time-independent nonlinear function [12]. In order to achieve a significance level of α=0.95 when confirming or rejecting each of the aforementioned three null hypotheses, we generate 19 surrogates for each of the three null hypotheses, after which original data and generated surrogates are compared by calculating the zeroth-order prediction error γ [12]. If the zeroth- order prediction error for the original recordings (γ_0) is smaller in comparison to the calculated error for surrogate data (ε), then a null hypothesis can be rejected, meaning that the analyzed time series does not originate from the assumed class of processes (i.e. starting null hypothesis).

For this purpose, surrogates are generated by using Matlab toolkit MATS, developed by Kugiumtzis and Tsimpiris [13], while γ is calculated in Matlab environment using script developed by Kaplan [14]. Surrogate data analysis is performed for time series both embedded into three-dimensional phase space with embedding delay τ=4 determined by the symplectic geometry method [15] and mutual information technique [16], respectively. Neighbors for prediction were sought amongst those points that were inside 5% of maximal distance to the referent point (i.e. point for which the prediction is made). In particular, points were selected from the range that is two and four times larger than the measurement error for Borca and Cuvarnica, respectively, in order to ensure large enough neighborhood for prediction (and, therefore, proper calculation of zeroth-order prediction error).

Regarding the first null hypothesis, it could be assumed solely by visually inspection of Figure 2 that the observed time series do not represent an example of random data. In order to test this claim, surrogates are generated by randomly shuffling the data without repetition, after which the zeroth-order prediction error is calculated both for the original recording and for the generated surrogates. Indeed, as it was already assumed, γ_0 is smaller than γ for all the examined cases, allowing us to reject the null hypothesis with significance level α=1 (Figure 3a).

For the purpose of testing the second null hypothesis, we employ the phase randomization analysis by randomizing Fourier surrogates of the original data and then by computing the inverse transform to obtain randomized time series [17]. In contrast to the previous case, prediction error for time series recorded at Borca station falls within the prediction error of more than a single surrogate, which further implies that we cannot reject the null hypothesis. On the other hand, $\gamma_0 > \gamma$ for only one surrogate (and not for all prediction steps) at Cuvarnica station, indicating that we could reject the null hypothesis with significance level α=0.95 (Figure 3b).

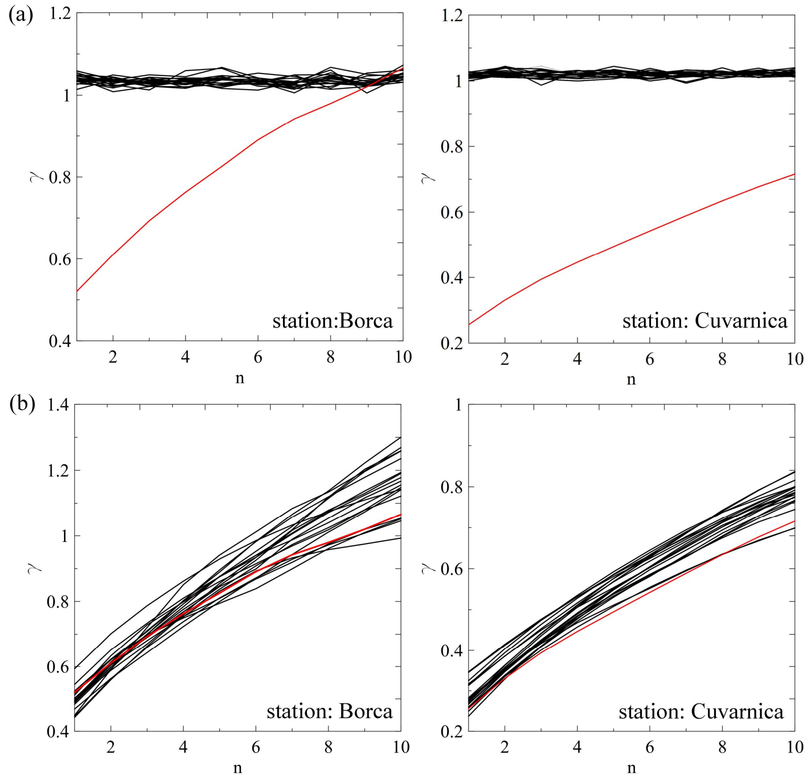


Fig. 3. Testing the first (a) and the second (b) null hypothesis. Red line denotes the zeroth-order prediction for the original time series and black lines - zeroth-order prediction for the surrogates. We could reject the first null hypothesis ($\gamma_0 < \gamma$) for both stations, as well as the second null hypothesis for recorded data at Cuvarnica station. For the observed time series at Borca station, we could not reject the second null hypothesis, since $\gamma_0 > \gamma$ for more than one surrogate.

Since the second null hypothesis was rejected for the observed groundwater level oscillation at Cuvarnica station, we also need to test the third hypothesis, in order to examine whether the recorded data originate from a stationary Gaussian linear process that has been distorted by a monotonic, instantaneous, time-independent nonlinear function. For this purpose, we calculate the amplitude adjusted Fourier-transformed (AAFT) surrogates [18], which is performed by rescaling the original data to a normal distribution, after which a Fourier-transformed surrogate of the rescaled data is constructed. The final surrogate is then scaled to the distribution of the original data. As apparent from Figure 4a, the null hypothesis could be rejected, since $\gamma_0 < \gamma$ for all the examined surrogates, with significance level α=1. However, generation of amplitude adjusted surrogates results in changes to the power spectrum of the final surrogate, further causing the power spectrum whitening of the original data, which could lead to unreliable results and assumptions. In order to improve the obtained results, a method of iterated AAFT surrogates (IAAFT) is applied by performing a series of iterations in which the power spectrum of AAFT surrogate is adjusted back to that of the original time series before the distribution is rescaled back to the original distribution. This is obtained by adjusting back the amplitudes of the Fourier transformed AAFT surrogates to the Fourier transformed surrogates of the rescaled original data. The obtained surrogates are then inverse Fourier transformed and rescaled back to the original data distribution by sorting the original data according to the ranking of the Fourier-transformed surrogate. These two steps are iterated for several times (in our case 500), until the whitening of the power spectrum becomes sufficiently small. Since the obtained results indicate that $\gamma_0 > \gamma$ for more than one surrogate, the third null hypothesis could not be rejected in this case.

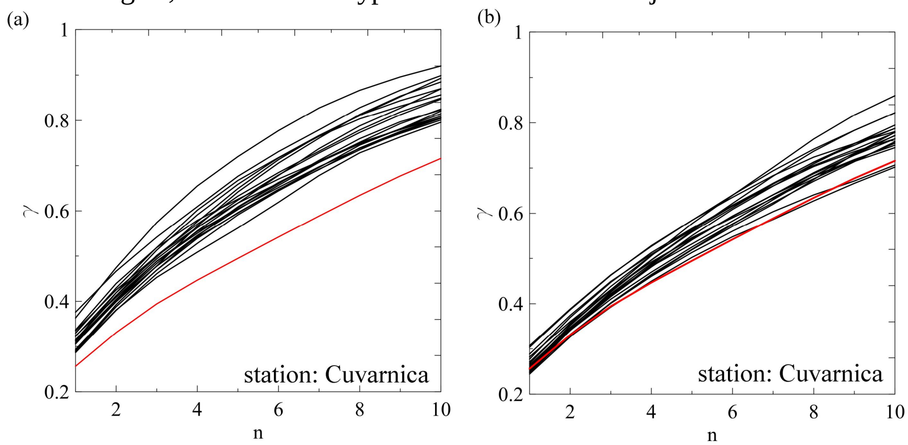


Fig. 4. Testing the third null hypothesis: (a) AAFT method; (b) IAAFT method. Red line denotes the zeroth-order prediction for the original time series and black lines - zeroth-order prediction for the surrogates. AAFT method resulted in $\gamma_0 < \gamma$ for the original data and generated surrogates allowing us to reject the null hypothesis, while $\gamma_0 > \gamma$ with IAAFT method for more than one surrogate, indicating that we could not reject the third null hypothesis.

In order to verify the results obtained by surrogate data testing, we need to conduct the determinism test [19], which assumes that if a time series originates from a deterministic process, it can be described by a set of more or less complex first-order ordinary differential equations, whose solution could be presented in a form of a corresponding vector field. If the observed time series originated from a deterministic system, the obtained vector field should consist solely of vectors that have unit length, indicating the average length of all directional vectors κ to be equal to 1. If solutions in the phase space are to be unique, then the unit vectors inside each box may not cross, since that would violate the uniqueness condition at each crossing. In other words, if the system is deterministic, the average length of all directional vectors κ will be 1, while for a completely random system $\kappa \approx 0$. However, it should be emphasized that Kaplan and Glass [19] indicated that an average length of the vector of the deterministic system equals 1 only for the infinite long data set.

So as to be able to conduct the determinism test, we firstly need to embed the observed scalar series into the appropriate phase space via the embedding procedure, originally proposed by Takens [20]. The embedding is performed by using the open-source software developed by Kodba et al.[8].

As a first step of embedding, we need to determine the optimal embedding delay. According to [16], the value of τ that produces the first local minimum of mutual information should be used for phase portraits. The values of optimal embedding delay calculated by using the average mutual information method for all the observed time series are shown in Figure 5(a).

The next step in our analysis, after calculating the optimal value of embedding delay, is to determine the minimal required embedding dimension m in order to fully resolve the complex structure of the attractor. In this paper, we use the procedure suggested in [22] that identifies the number of ’’false nearest neighbors’’, points that appear to be nearest neighbors because the embedding space is too small. Generally, there are two criteria that are usually applied when using ‘’false nearest neighbor’’ method. We consider that normalized distance between the embedding coordinates of two presumably neighboring points is larger than a given threshold (R_{tr}), if these two point are false neighbors. The results obtained with the false nearest neighbor method for acceleration time histories at the observed stations are presented in Figure 5(b). It is clear that the percentage of false nearest neighbors decreases to value <<0.01 at embedding dimension m=3 in both cases.

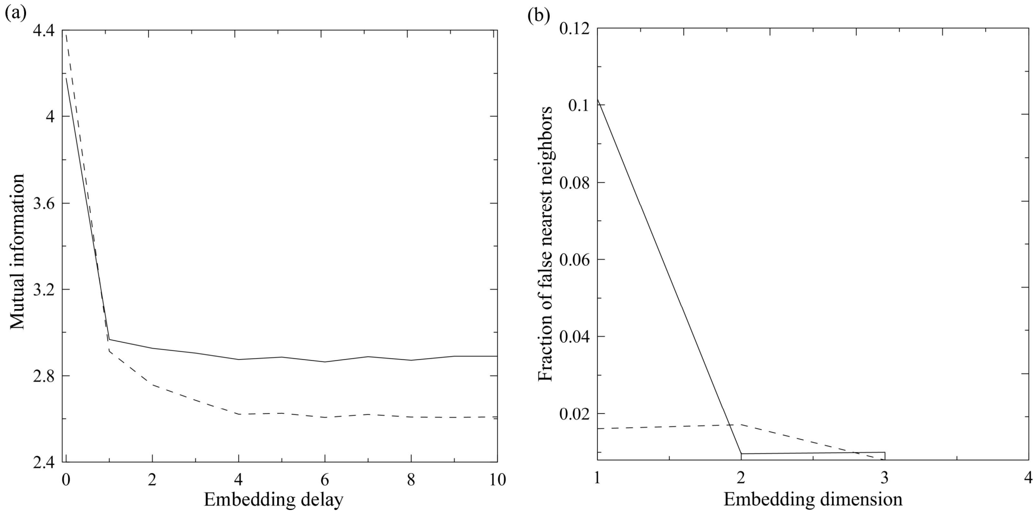


Figure 5. (a) Determination of the proper embedding delay for recorded groundwater dynamics at Borca (solid line) and Cuvarnica (dashed line) **-** the average mutual information has the first minimum at τ=4 in both cases; (b) Determination of the minimal required embedding dimension for recorded groundwater dynamics at Borca (solid line) and Cuvarnica (dashed line). It is clear that the fraction of false nearest neighbors decreases to value <<0.01 at m=3 in both cases.

In order to conduct the determinism test, we coarse grained the previously calculated embedding space into maximum 42^3 grid for data observed at both stations. For calculating the determinism factor, we included only those boxes visited at least one time by the trajectory. The obtained value of determinism factor κ is 0.78 for both stations, indicating that the level of stochasticity is still high as experimental recordings of deterministic signals usually have $\kappa > 0.9$ [12]. This is further confirmed by constructing the corresponding vector field composed of vectors of different sizes for both monitoring stations (Figure 6).

6. Conclusion

In present paper we analyzed the groundwater level dynamics in order to determine the main pattern of the background mechanism. For this purpose, we applied methods of nonlinear time series analysis, including surrogate data testing, delay embedding theorem and determinism test. As a case study, we chose observed recordings at two piezometers installed in Pancevacki rit hydrogeological province, at Borca and Cuvarnica station. Both piezometers are 12m deep, i.e. the observed groundwater dynamics belongs to the same aquifer. Recorded time series are analyzed for the period 2007-2013, since no or only few recordings are available for 2005 and 2006 (although permanent monitoring of groundwater dynamics is performed since 1994).

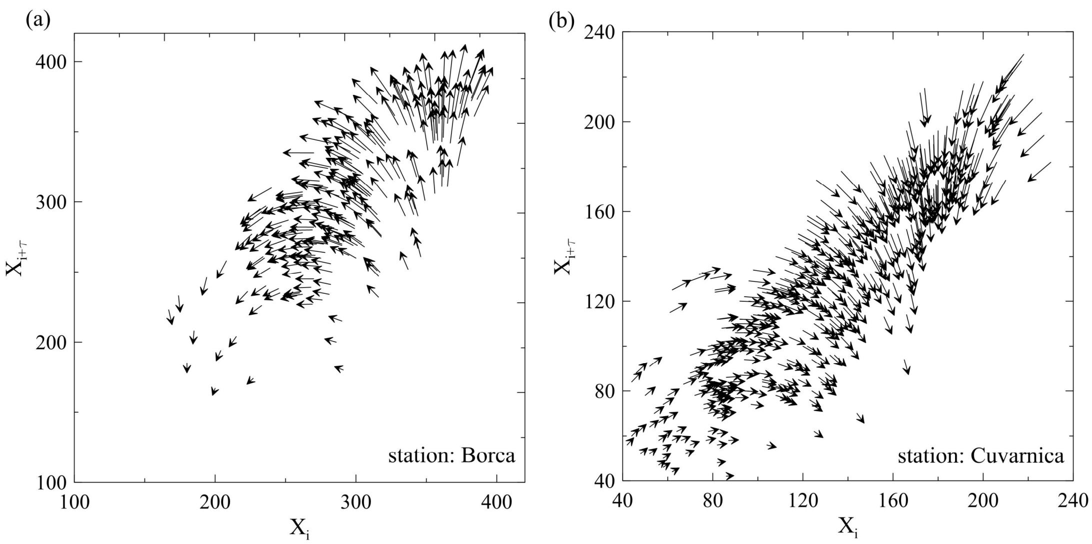


Fig. 6. Determinism test. The approximated vector field for the embedding space reconstructed with $\tau = 4$ and $m = 3$. The pertaining determinism factor is $\kappa = 0.78$ for the recorded groundwater level oscillation at (a) Borca and (b) Cuvarnica station.

The obtained results indicate that main mechanism in the background of both time series belong to the class of stochastic linear processes, with Gaussian inputs (Borca), which could be distorted by a monotonic, instantaneous, time-independent nonlinear function (Cuvarnica). Such findings are firstly implied by surrogate data testing, and further confirmed by determinism test, where rather low value of determinism factor (κ=0.78) implies the high level of stochasticity in the observed time series.

In conclusion, one should note that although the obtained results are convincing and confirmed by different methods, further analysis of groundwater dynamics should certainly include longer period of time and with the recordings observed at stations in different aquifers, i.e. various geological surroundings. In that way, our preliminary results could be further confirmed, making a way towards creating the appropriate prediction model, which should be the final goal of the present analysis.

Acknowledgements

This work is partly supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia, Contracts No.37005, 171017 and 35046.

References

[1] Adamowski, J., Chan, H.F., *A wavelet neural network conjuction model for groundwater level forecasting*, Journal of Hydrology, Vol. 4074, 28-40, 2011.

[2] Affandi, A.K., Watanabe, K., Tirtomihardjo, H., *Application of an artificial neural network to estimate groundwater level fluctuation*, Journal of Spatial Hydrology, Vol. 7, 23-46, 2007.

[3] Daliakopoulos, I.N., Coulibaly, P., Tsanis, I.K., *Groundwater level forecasting using artificial neural networks*, Journal of Hydrology, 1-12, 2005.

[4] Guistolisi, O., Simeone, V., *Optimal design of artificial neural networks by a multiobjective strategy: groundwater level predictions*, Hydrological sciences-Journak-des Sciences Hydrologiques, Vol. 51, 502-523, 2006.

[5] Nayak, P.C., Satyaji Rao, Y.R., Sudheer, K.P., *Groundwater level forecasting in a shallow aquifer using artificial neural network approach*, Water Resources Management, Vol. 20, 77-90, 2006.
[6] Shirmohammadi, B., Vafakhah, M., Moosavi, V., Moghaddamnia, A., *Application of several data-driven techniques for predicting groundwater level*, Water Resources Management, Vol. 27, 419-432, 2013.
[7] http://www.hidmet.gov.rs/podaci/hidro_pod_godisnjaci/
[8] Kodba, S., Perc, M., Marhl, M., *Detecting chaos from a time series*, European Journal of Physics, Vol. 26, 205-215, 2005.
[9] Kostić, S., Vasović, N., Perc, M., Toljić, M., Nikolić, D., *Stochastic nature of earthquake ground motion*, Physica A: Statistical Mechanics and its Applications, Vol. 392, 4134-4145, 2013a.
[10] Kostić, S., Perc, M., Vasović, N., Trajković, S., *Predictions of experimentally observed stochastic ground vibrations induced by blasting*, *PLoS One* 8, e82056, 2013b.
[11] Milojevic, N., Filipovic, B., Dimitrijevic, N., *Hydrogeology of the territory of Belgrade city*, Obod, Cetinje, 1975 (in Serbian).
[12] Perc, M., Green, A.K., Jane Dixon, C., Marhl, M., *Establishing the stochastic nature of intracellular calcium oscillations from experimental data*, Biophysical Chemistry, Vol. 132, 33-38, 2008.
[13] Kugiumtzis, D., Tsimpiris, A., *Measures of analysis of time series (MATS): a MATLAB toolkit for computation of multiple measures on time series data bases*, Journal of Statistical Software, Vol. 33, 1–30, 2010.
[14] http://www.macalester.edu/~kaplan/, 2015
[15] Lei, M., Wang, Z., Feng, Z. 2002. A method of embedding dimension estimation based on symplectic geometry. Phys. Lett. A, 303, 179-189.
[16] Fraser, A., Swinney, H., *Independent coordinates for strange attractors from mutual information*, Physical Review A, Vol. 33, 1134-1140, 1986.
[17] Kantz, H., Schreiber, T. *Nonlinear time series analysis*, 2nd ed., Cambridge University Press, Cambridge, 2004.
[18] Dolan, K.T., Spano, M.L., *Surrogate for nonlinear time series analysis*, Physical Review E, Vol. 64, 046128-1-6, 2001.
[19] Kaplan, D.T., Glass, L., *Direct test for determinism in a time series*, Physical Review Letters, Vol. 68, 427–430, 1992.
[20] Takens, F. *Detecting strange attractors in turbulence*, in: D.A. Rand, L.S. Young (Eds.), Dynamical Systems and Turbulence, in: Lecture Notes inMathematics, vol. 898, Springer-Verlag, pp. 366–381, 1981.
[21] Kennel, M., Brown, R., Abarbanel, H. *Determining embedding dimension for phase–space reconstruction using a geometrical construction*, Physical Review A, 45, 3403–3411, 1992.