International Journal of Earth, Energy and Environmental Sciences
ISSN: 2517-942X
Vol:5, No:5, 2011

2D Graphical Analysis of Wastewater Influent Capacity Time Series

Monika Chuchro, Maciej Dwornik

***Abstract*—**The extraction of meaningful information from image could be an alternative method for time series analysis. In this paper, we propose a graphical analysis of time series grouped into table with adjusted colour scale for numerical values. The advantages of this method are also discussed. The proposed method is easy to understand and is flexible to implement the standard methods of pattern recognition and verification, especially for noisy environmental data.

***Keywords*—**graphical analysis, time series, seasonality, noisy environmental data

I. Introduction

THE concept of time series graphical analysis, a new complementary tool for trend and seasonal dependency analysis is present in this article. This method can be also useful for assessment of external factors impact on the analyzed variable. During the past decade various structure and pattern recognition approaches were added to the arsenal of available diagnostics tools: Golyandina et al. present a singular spectrum analysis and related techniques for structure recognition and change-point detection [1]. Kitagawa used genetic algorithms for patters exude [2]. Modeling of influent capacity by artificial neural networks for structure recognition and prediction discussed by Gamal El-Din [3]. Markov chain as a tool for analysis of causality and for visualization of graphical models [4] and automated image analysis present by Hu et al. [5]. Mathematical models for time series structure are well develop. They are, however, work under assumption of prior knowledge about distribution and preliminary information of data structure. Presented method need not any previous assumption. Graphical analysis could be a good tool for preliminary analysis of time series.

II. Data

The data we analyze consist of wastewater treatment plant daily influent capacity in Km^3/day in eight consecutive years from 2000 to 2007. The wastewater treatment plant is the biggest object localized in central Poland, with average daily influent capacity near to 180 $Km^3 day^{-1}$ and median 175 $Km^3 day^{-1}$. Analyzed data are stationary, skewed right with value 0.91 and leptokurtic. A distribution of data with skewness and leptokurtic is seen in Figure 1. Sewage system collecting wastewater from household and industry also collect stormwater runoff. It is a reason why data were checked for correlation between daily precipitation amount and influent time series. Pearson correlation coefficient of those two variables equals 0.1, Kendal tau correlation is higher and equals 0.28, the highest result of relationship between variables is for Spearman’s rank correlation (0.38). Results of correlation coefficients indicate nonlinear influence of precipitation on influent capacity.

Influent capacity data were prepared for analysis. Gaps in time series were filled up with moving 1D averaged method. Incidental errors were replaced by weighted mean of neighbourhood’s data. Data were reorganized into table with 7, 8, 14, 28 rows contained influent capacity values, connected with colour scale. Vertical axis indicates days in the figure 2 to figure 4. In the first line in the graph are situated Saturdays, in the first column were data from Saturday 1st to 28th of January 2000 year. In the figure 3 on the horizontal axis are situated days of the year, on the vertical axis years from 2000 mark as 1 to 2007 mark as 8. From accomplished images noise were eliminated before patterns recognition. Several filters were tested: average filter, median filter and adaptive filter (Wiener filter). Filtration were made in 3x3 pixel filter mask. Results of filtration was shown in fig.2-4.

Average filtration is classical low-pass filtration, which assign central point image under mask average of area covered by mask. Missing data at the boundary of image are assumed by mirror-reflecting. This is very fast and easy filter. Main disadvantage is blurring edge and propagate extremely value to neighboured pixels. One of the method to solve this problem is adaptive filtration, which choose power of blurring to variance in area covered by mask. Power of filtration is inverse to variance.

Median filtrarion is most powerful method to eliminate salt & pepper noise type. Extremely value are replaced by median of values from area covered by mask.

The well know median filters [6] were use. Low-pass averaging 2D filers was also tested. To minimalize the patterns damage various filters were compared. The best results for removing outlier values, but leaving patterns have median filter in window 3x3 with option symmetric.

Monika Chuchro is with the AGH University of Science and Technology, al.Mickiewicza 30, Cracov 30-059, Poland (e-mail: chuchro@ geol.agh.edu.pl).
Maciej Dwornik is with the AGH University of Science and Technology, al.Mickiewicza 30, Cracov 30-059, Poland (e-mail: maciek.dwornik@wp.pl).

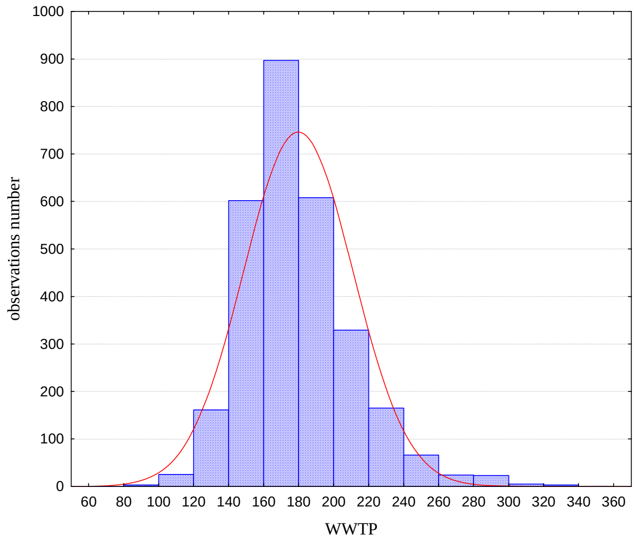

Fig. 1 Influent capacity time series histogram

III. Analysis

As it can be seen, in the Fig.2, 3 and 4 the evident lineaments appear. All visible structures seen in the Figure 2 have vertical shape. Lineaments of high influent capacity usually have sharpen edges. Whereas low influent capacity structures are broader with fuzzy edges. The time series images were compared with feasts days in Poland. In the pictures were marked: New Years, Easters, 1 and 3 of May, All's Souls Day, 11th of November and Christmas. Patterns were also compared with the weather data, especially with precipitation. An average of precipitation amount during the analyzed years is changing. The 2000 year had the most intensive precipitation with average 2.3 and maximum 151 mm on 310th day of. First three years and the last one (2007 year) have precipitation average above 1.6 mm. The most dry year among analyzed was 2005 year with average precipitation equal 1.3 mm and 216 days without precipitation.

IV. Results

Time series image seen in Figure 2 have a seasonal structure. The picture could be divided in to eight, yearlong pieces. The first three years have higher influent capacity. Next three have low influent capacity. At the end of the analyzed period of time influent capacity arose once more.

There is a possibility that analyzed time series data have systematic long-term pattern, years with higher influent capacity are interlace with years with lower influent capacity. The long-term pattern in analyzed data should be connected with long-term patterns of precipitation data. Also it should be connected with economy, water and wastewater prices, cities growth.

In the figure 3 we can see a whole year structure with a one day shift cause by 29th of February. This shift happen in year: 2000- line first and 2004- line fifth. This shift is visible in feast days lineaments, especially in Christmas. The lines of Christmas in 2000 and 2004 are moved, it is seen in Figure 3. There is no one well-defined year structure of influent capacity. One common pattern is influent capacity decreasing occur on months: May, June and July.

Furthermore different filters mentioned above were tested on influent time series. Averages filtration cause decreasing contrast between seasons (Fig.3). Wiener's filter have larger span of data than median and averaging filter. The horizontal continuity is better visible on averages filtration, but the vertical on median filtration. The adaptive Wiener filtration did not prove with influent data.

Small number of dispersed red dots, combined into linear groups 4 to 8 days long may be connected with intensive precipitation. What is interesting, incidentals with high influent capacity happen more frequently in the second and third month of the year (Fig.4).

Comparison of feast days with time series images shows influent capacity decreasing during feast days and visible increasing of influent, a week length, before the main feast days: Easters and Christmases. The linear structure localized in the right side of figure 2 from 98 column 6 line to 103 column 18 line is accidental cause by rain. Days number: 2780, 2782, 2809, 2810, 2870, 2871 were days with precipitation higher than 6 mm. During the adjacent days precipitation were also observed.

Multiple regression, Census II variant X11 and Fourier analysis confirm detected in image analysis patterns. Statistically significant were binary variables describing seasonality and lineaments: March, July, December, feast days, week before Easters and Christmas. Also regression model contain binary variable of: Wednesdays, Saturdays and Sundays, precipitation, humidity, average temperature, influent capacity lag 1 and 365.

Trend in data is not statistically significant. Probably if analyzed would be repeat on the wilder set of data, sinusoid parabola trend would be approvable.

Those methods also recognize a week influent structure with increasing capacity on Wednesdays and Saturday and decreasing capacity on Sundays. Those lineaments were not visible in filtrated data.

V. Concluding Remarks

Graphical analysis can be useful tool for time series structure recognition. Lineaments and patterns are visible and easy to identify in the colour image. Proposed analysis helps to chose the proper methods of further analysis and tests. The biggest problem could be only with correct data filtration. Finding appropriate method of filtration require carry out many trials. Presented graphical analysis of time series will be extend by Hough transform and automatic pattern recognition [7].

ACKNOWLEDGMENT

This work has been financially supported by Faculty of Geology, Geophysics and Environmental Protection, AGH University of Science and Technology, grant no 18.18.140.942

REFERENCES

[1] Golyandina N., Nekrutkin V., Zhigljavsku A.Z, Analysis of Time Series Structure; SSA and related techniques, Chapman &Hall/CRC Monographs on Statistic & Applied Probability, Boca Raton 2001.

[2] Kitagawa G., Introduction to Time Serieas Modeling, Chapman &Hall/CRC Monographs on Statistic & Applied Probability, Boca Raton 2009.

[3] Gamal El-Din A., Smith D.W., “Modelling approach for high flow rate in wastewater treatment operation”, Journal of Environmental Engineering and Science 2000, vol. 1, pp. 275-291.

[4] Dahlaus R., Eichler M., “Causality and graphical models in time series analysis”, http://galton.uchicago.edu/~eichler/hsss.pdf.

[5] Hu Y., Osuna-Highley E., Hua J., Nowicki T.S., Stolz R., McKayle C., Murphy R.F., “Automated analysis of protein subcellular location in time series images”, Bioinformatic 2010 vol.26 no. 13, pp.1630-1636.

[6] Zvornarev P.S., Apalkov I.V., Khryashchev, Priorov A.L., “Adaptive Switching Median Filter with Neural Network Impulse Detection Step”, Lecture Notes in Computer Science, 2005 Volume 3696/2005 pp. 537-542.

[7] Rong-Chin L., Wen-Hsiang T., “Gray-Scale Hough transform for thick line detection in gray-scale images”, Pattern Recognition,1995, vol. 28, no.5, pp.647-661.

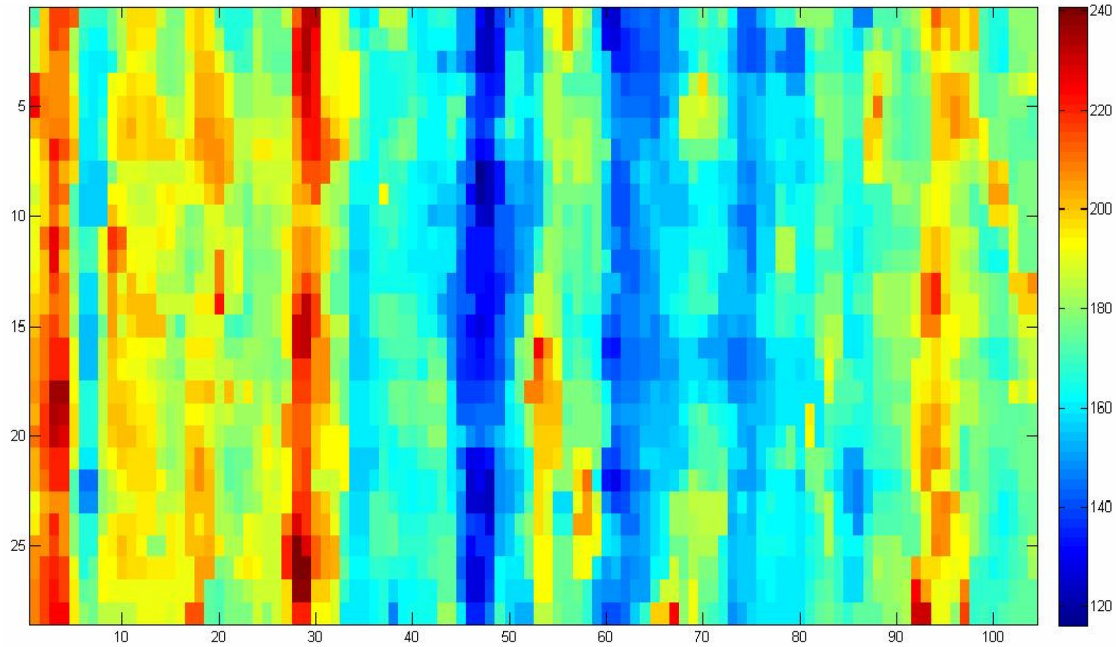

Fig.2 Four- weeks time series after median filter [3x3]

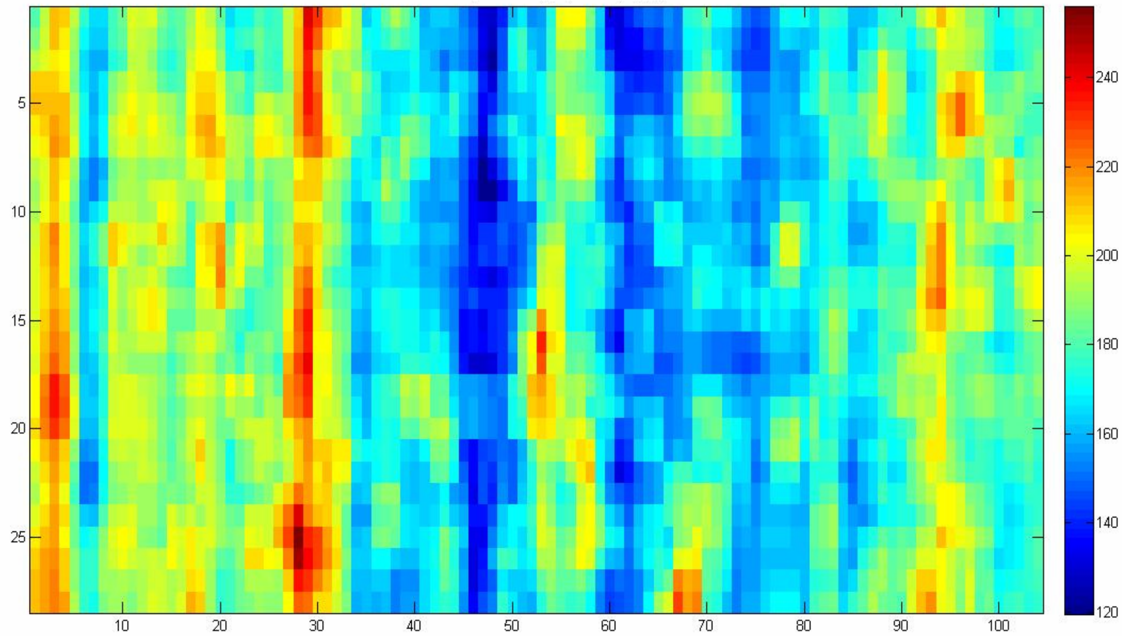

Fig.3 Four- weeks time series after averaging filter [3x3]

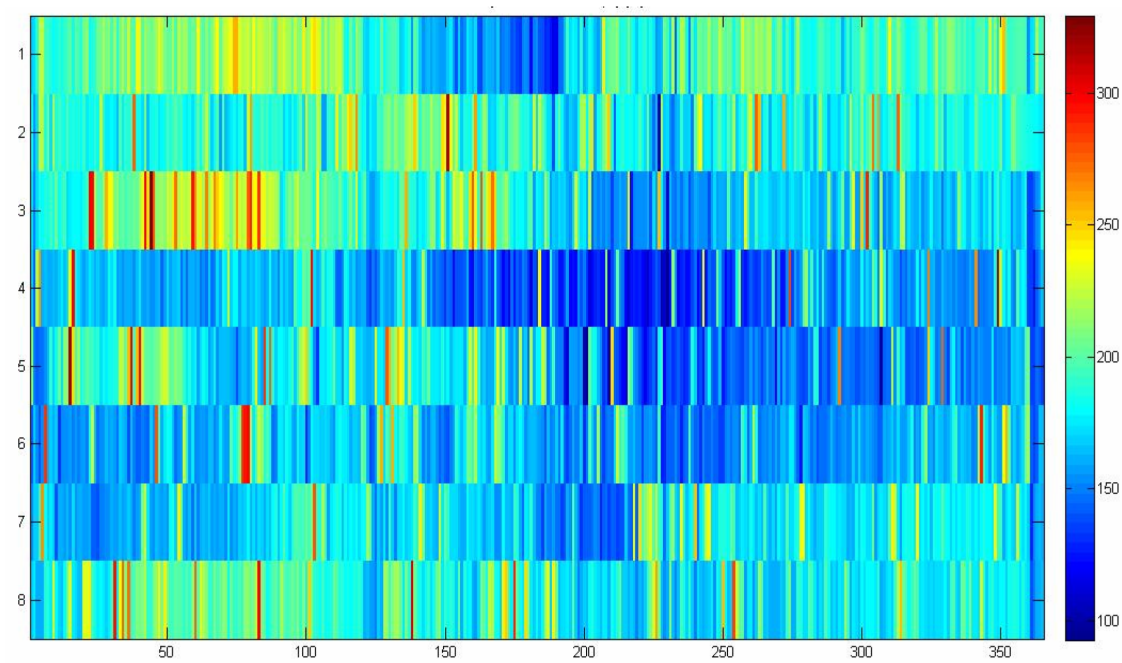

Fig. 4 An years time series after median filter [3x3]