Up and Down Trend Associations in Analysis of Time Series Shape Association Patterns

Ildar Batyrshin

Mexican Petroleum Institute, D.F., Mexico
batyr1@gmail.com

Abstract. The method of recognition of shape association patterns with direct and inverse relationships is proposed. This method is based on a new time series shape association measure based on Up and Down trend associations. The application of this technique to analysis of associations between well production data in petroleum reservoirs is discussed.

Keywords: Time series shape association measure, time series pattern, oilfield.

1   Introduction

Last years it grows the interest in development of time series data mining models and methods [8-15]. One of the important components of time series data mining methods is a measure of time series similarity or association that can be used for clustering time series, searching similar time series and time series patterns [1,2,8,9,13]. Many such measures have been introduced and studied [1-3,8,10,14,15,]. In [4] it was proposed to use axiomatic approach to analysis of classes of time series shape similarity and shape association measures. The set of axioms introduced in [4] defines two classes of such measures: 1) the class of shape similarity measures that can be used for measuring possible direct relationships between time series or time series patterns, and 2) the class of shape association measures that can measure both direct and inverse relationships between them. Most of the known time series similarity measures cannot measure inverse relationships between time series and belong only to the class of similarity measures. But there are many economical, financial, industrial, ecological systems that contain changing in time elements or characteristics such that an increase in the values of the one of them happens together with a decrease in the values of another one: prices and sales, the sales volumes of competitive companies, the wind velocity and air pollution concentration etc. In [4] it was pointed out that correlation coefficient and trend association measure based on moving approximation transform [3] satisfy axioms of shape association measures and can be used for measuring possible direct and inverse relationships between time series and time series patterns.

In presented work a new measure of time series shape associations is introduced that can be used for recognition of time series shape association patterns both with direct and inverse possible relationships. This measure uses only qualitative information

about increasing (Up) and decreasing (Down) time series values. This measure uses only the slope sign but not the slope value. In many applications it is more reliable information about the change of time series values. Moreover, in comparing dynamics of two time series often only qualitative (Up and Down) information about the change of time series values can be valid if the possible associations between time series are non-direct and may be caused by some hidden factor or by group of unknown factors.

The paper is organized as follows. The classes of time series shape similarity and shape association measures are considered in Sections 2. In Section 3 a new time series shape association measure is introduced. In Section 4 the method of recognition of Up and Down shape association patterns is proposed. In Section 5 an example of application of this method in petroleum industry is discussed. The last section contains conclusions.

2   Shape Similarity Measures

A time series of length n, $(n>1)$, is a sequence of a real values $y = (y_1,…,y_n)$ given at time points $t= (1, …,n)$. Denote T^n the set of all time series of length n. Suppose p,q $(p \neq 0)$ are real values. Denote $py+q = (py_1+q, …,py_n+q)$.

Definition 1[4]. A time series **shape similarity measure** is a function $S:T^n \times T^n \rightarrow [0,1]$ satisfying the following axioms:

P1. $S(x,y) = S(y,x)$,   (simmetry),
P2. $S(y,y) = 1$,   (reflexivity),
P3. $S(y+q,x) = S(y,x)$, for any $q \geq 0$,   (translation invariance).

A shape similarity measure satisfying the axiom:

P4. $S(py,x) = S(y,x)$, if $p > 0$,   (scale invariance),

is called a **scale invariant shape similarity measure.**

Definition 2 [4]. A time series **shape association measure** is a function $S:T^n \times T^n \rightarrow [-1,1]$ satisfying axioms P1-P4 and

P5. $S(-x,y) = -S(x,y)$,   (inverse relationship).

Consider examples of time series shape similarity and shape association measures [4]. Suppose D is a distance measure defined on T^n and $F_k(y)=(F_k(y_1),… F_k(y_n))$, $k=1,2$, is one of the following time series normalization functions defined for all $i=1,2,…,n$ as follows:

$$F_1(y_i) = y_i - \bar{y}, \text{ where } \bar{y} = \frac{1}{n}\sum_{j=1}^{n} y_j,$$

$$F_2(y_i) = \frac{y_i - y_{\min}}{y_{\max} - y_{\min}}, \text{ where } y_{\min} = \min\{y_1,...,y_n\}, \quad y_{\max} = \max\{y_1,...,y_n\}.$$

Define a function $S:T^n \times T^n \to [0,1]$ as follows:

$$S_{D,F}(x,y) = 1 - \frac{D(F(x),F(y))}{M} \quad,$$

where M is a constant such that $M \geq D(F(x),F(y))$ for all x,y in T^n. It can be shown [4] that $S_{D,F}$ is a shape similarity measure if $F = F_1$ and S is a scale invariant shape similarity measure if $F = F_2$.

The correlation coefficient [7] gives an example of time series shape association measure satisfying axioms P1-P5.

Another example of time series shape association measure gives the trend association measure based on moving approximation transform [3]. Consider time series from T^n. A window W_i of size $k \in \{2,...,n\}$ is defined as a sequence of indexes $W_i = (i,i+1,...,i+k-1)$, $i \in \{1,..., n-k+1\}$. A sequence $J_k = (W_1, W_2,..., W_{n-k+1})$ of all windows of size k is called a sliding window of size k. Note that if $k = n$ then $J_n = (W_1)$ and we have only one window containing all indexes $(1,2,...,n)$. Suppose y is a time series in T^n and J_k is a sliding window of size k, $k \in \{2,...,n\}$. Denote $y_{W_i} = (y_i, y_{i+1},...,y_{i+k-1})$ the values of y in time points $(i,i+1,...,i+k-1)$ defined by window $W_i = (i,i+1,...,i+k-1)$. Linear functions $f_i = a_i t + b_i$, $(i=1,...,n-k+1)$, with parameters $\{a_i,b_i\}$ minimizing the criterion

$$Q(f_i, y_{W_i}) = \sum_{j=i}^{i+k-1} \left(a_i j + b_i - y_j\right)^2 \,,$$

are called moving (least squares) approximations of y_{Wi}. A sequence $MAP_k(y) = (a_1, ..., a_{n-k+1})$ of slope values of moving approximations of time series y in a sliding window of size k is called a moving approximation (MAP) transform of time series y. The slope values a_i are called local trends. The values of MAP transform $MAP_k(y)$, $k \in \{2,...,n\}$, can be calculated as follows [3]:

$$a_i = \frac{6 \sum_{j=0}^{k-1} (2j - k + 1) y_{i+j}}{k(k^2 - 1)} \,, \; i \in \{1,..., n-k+1\}. \tag{1}$$

Suppose $y = (y_1,...,y_n)$, $x = (x_1,...,x_n)$ are two time series and $MAP_k(y) = (a_{y1}, ..., a_{ym})$, $MAP_k(x) = (a_{x1}, ..., a_{xm})$, $(k \in \{2,...,n-1\}$, $m = n - k+1)$, are their MAP transforms. The following function is called a measure of trend associations:

$$coss_k(y,x) = \frac{\sum_{i=1}^{m} a_{yi} \cdot a_{xi}}{\sqrt{\sum_{i=1}^{m} a_{yi}^2 \cdot \sum_{j=1}^{m} a_{xj}^2}} \,. \tag{2}$$

It can be shown [4] that a measure of trend associations is a time series shape association measure, i.e. satisfies axioms P1-P5.

3   Up and Down Trend Associations

Define Up and Down trend association measure as follows. In MAP transform and in measure of trend associations we replace trend values a_i by their sign values

$$A_i = sign(a_i). \tag{3}$$

These values will be called Up and Down trends, or for short UD-trends, and MAP transform for given time series y will be replaced by a sequence of UD-trends:

$$UDT_k(y) = (A_{y1}, …, A_{ym}). \tag{4}$$

From (1) we obtain:

$$A_i = sign(\sum_{j=0}^{k-1}(2j-k+1)y_{i+j}), \tag{5}$$

and (2) will be replaced by

$$S_{UDk}(y,x) = \frac{\sum_{i=1}^{m} A_{yi} \cdot A_{xi}}{\sqrt{\sum_{i=1}^{m} A_{yi}^2 \cdot \sum_{j=1}^{m} A_{xj}^2}}. \tag{6}$$

Note that in (3)-(6) all A_i take the values from the set $\{-1,0,1\}$.

Proposition 1. Up and Down trend association measure (6) for any $k \in \{2,…,n\}$ is a shape association measure, i.e. satisfies axioms P1-P5.

Below we will consider the simplest case when $k=2$. In this case S_{UD2} will be denoted as S_{UD}, slope values (5) will be calculated as follows:

$$A_i = sign(y_{i+1}-y_i), \tag{7}$$

and Up and Down trend sequence (4) will have the following form:

$$UDT(y) = (A_{y1}, …, A_{y(n-1)}). \tag{8}$$

Suppose g is a strictly increasing monotonic function on the set of real numbers, i.e. from $u < v$ it follows $g(u) < g(v)$. Define a strictly monotonic transformation G on the set of time series T^n as follows:

$$G(y) = (g(y_1),…, g(y_n)).$$

Proposition 2. A time series shape association measure $S = S_{UD}$ is invariant under strictly monotonic transformations, i.e. satisfies the property:

P6. $S(G(y),x) = S(y,x)$,   if G is a strictly monotonic transformation.

Consider synthetic example of two time series x and y given in Table 1 and presented in Fig. 1. It is clear that these time series have inverse relationship because when one time series values increase the values of another one decrease and vice versa. Up and Down trend association measure can find inverse relationship between these time series: S_{UD}= -1, as we expected. It should be noted that correlation coefficient [7] satisfies axioms P1-P5 but cannot detect inverse relationships between these time series because correlation between these time series equals 0.

Table 1. Example of two synthetic time series

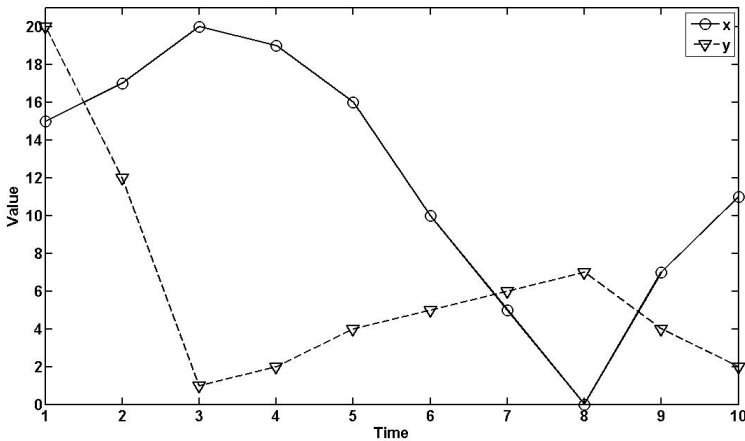| I	1	2	3	4	5	6	7	8	9	10
x	15	17	20	19	16	10	5	0	7	11
y	20	12	1	2	4	5	6	7	4	2

Fig. 1. Example of two synthetic time series

4   Up and Down Time Series Shape Association Patterns

Define for any time series x,y in T^n a time series shape association string as element-wise multiplication of UD-trends $UDT(y) = (A_{y1}, \ldots, A_{y(n-1)})$ and $UDT(x) = (A_{x1}, \ldots, A_{x(n-1)})$ obtained for window size k=2:

$$sas(x,y) = UDT(x) \cdot UDT(y) = (A_{y1}A_{x1}, \ldots, A_{y(n-1)}A_{x(n-1)})=(C_{xy1,\ldots,}C_{xy(n-1)}), \qquad (9)$$

where $C_{xyi}=A_{yi}A_{xi}$ for all $i = 1,\ldots,n$-1. Note that $C_{xyi} = 1$ if A_{yi} and A_{xi} have the same sign, i.e. $A_{yi}= A_{xi} =1$ or $A_{yi}= A_{xi} =$-1. In such cases we will say that A_{yi} and A_{xi} have positive association. We have $C_{xyi} = $-1 if A_{yi} and A_{xi} have opposite signs. In this case we will say that A_{yi} and A_{xi} have negative association. For completeness we can introduce 0-association if $A_{yi}=0$ or $A_{xi} = 0$. Depending on applications 0-associations can be considered as "positive" or as "negative" associations.

Definition 3. Positively (negatively, 0-) associated patterns in time series x and y are sequences $(x_k,…,x_{k+m})$ and $(y_k,…,y_{k+m})$, $(k,m \geq 1$ and $k+m \leq n)$ defined by the maximal sequences of UD-trends in $UDT(x)$ and $UDT(y)$ that have positive (negative, 0-) associations. For such sequences a number $m+1$ is called a length of the pattern.

From definition of positively associated patterns it follows that $C_{xyi}=A_{yi}A_{xi}>0$ for all i in $\{k,…,k+m-1\}$ and $C_{xyi}= A_{yi}A_{xi}\leq 0$, if $i= k-1$ or $i=k+m$. Inverse relations we have for negatively associated patterns. Any two time series from T^n can be represented as sequences of positively, negatively and 0- associated patterns with different length. In extreme case these time series can consist of one pattern. In applications it is usually have sense to find sufficiently large associated patterns in time series with the length greater than some heuristically defined number.

An algorithm of recognition of positively, negatively and 0- associated patterns follows from their definition. For given time series x and y calculate UD-trend sequences by (7) and (8), then calculate shape association string $sas (x,y)$ by (9) and find in this string all maximal substrings containing 1, all maximal substrings containing -1 and all maximal substrings containing 0.

5   Example of Recognition of Associated Patterns in Wells Production Data

The method of recognition of positively and negatively associated patterns in time series data bases described in previous section was implemented in VMD-Petro® (see [5]), a visualization and data mining tool for analysis of oilfield data. VMD-Petro® helps to petroleum engineers and researchers in analysis of reservoir dynamics based on history data and in monitoring of different oilfield parameters changing in time. The main goal of analysis of reservoir dynamics is to propose the methods and treatments that will increase or maintain on some level the oil recovery. The available information about reservoir dynamics is usually given by various measurements made in reservoir wells. The volume and the quality of this information depend on traditions, technological level of exploitation and measurements used in current field, on complexity of reservoir, on its productivity etc. The most available information about dynamics of reservoir is monthly oil, water and gas production data. In some fields it is used daily monitoring of gas injection volumes in injection wells or daily measurements (chromatography) of gas components in producing wells. Pressure measurements in wells can be done several times in month or only several times in year. The same situation takes place usually with water analysis and other types of data. High productive fields can be better equipped and have more data. As usually for many old oilfields it is available less data for the first years of exploitation than for the last years. Except of the problem of availability of data describing reservoir dynamics there is a problem of a quality of data. Usually they contain errors arisen in the process of measurement, rewriting and digitalization of data. Another problem is that often production data of wells are obtained as a result of measurement of production volume in a battery of production wells and further splitting of the measured value between wells joined in the battery. Such procedure of calculation of well production data usually contains their own errors.
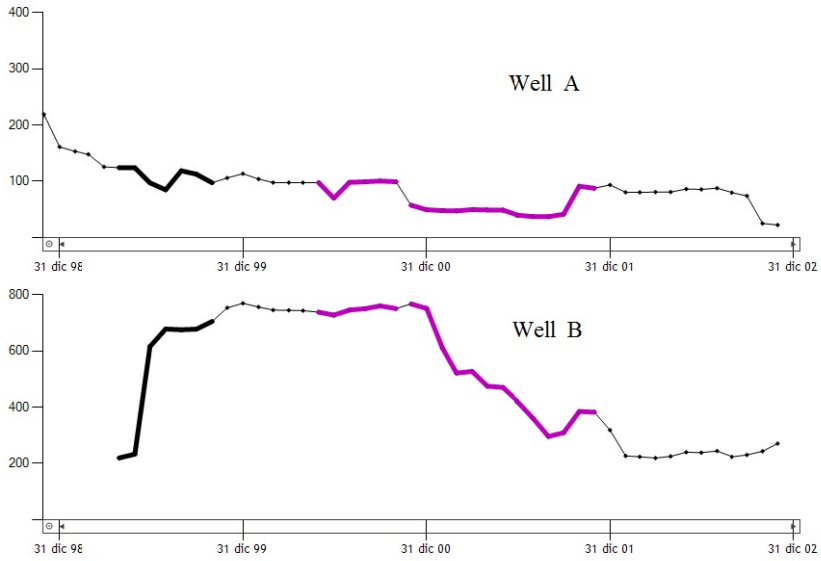
Fig. 2. Negatively associated (bold, left) and positively associated (bold, right) patterns of two time series of oil production in wells A and B
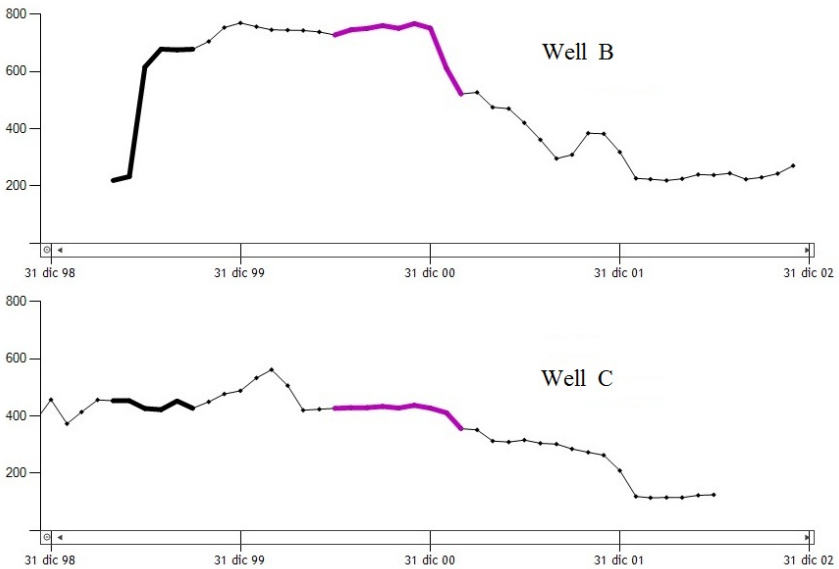

Fig. 3. Negatively associated (bold, left) and positively associated (bold, right) patterns of two time series of oil production in wells B and C

For the reasons described above the validity of data available for analysis does not high and qualitative Up and Down trend information about dynamics of wells production is more valid than quantitative information about values of production. VMD-Petro® has several modules for analysis and processing time series of oil, water and gas production. The data mining module gives possibility to find positively and negatively associated patterns in pairs of time series. The presence of such patterns in neighboring wells can be used for generation hypothesis about possible interference between them or about existing of some common mechanism depending, for example, on reservoir gas or water pressure, that influences on such association between wells production data. Figures 2 and 3 present a screen shots of VMD-Petro data mining module that shows large negatively associated (bold, left) and positively associated (bold, right) pairs of patterns of oil production data in different wells of some oilfield of Mexico. The presence of such patterns gives possibility to generate hypothesis about possible interference between wells. For example, negatively associated patterns between wells B and A and between B and C during the first months of exploatation of well B can give reasons for generation of hypothesis about possible interference between well B and his neighboring wells A and C.

6   Conclusions

In this paper a new approach to analysis of direct and inverse associations between time series and time series patterns is proposed. This approach uses a qualitative Up and Down trend information about time series dynamics that in many applications is more valid than information about time series values. The proposed methods of recognition of positively and negatively associated patterns in time series have been implemented in VMD-Petro®, a visualization and data mining software tool for analysis of oilfield data, developed in Mexican Petroleum Institute. The proposed method can be applied also in analysis of associations between time series in various application areas: in economics, finance, ecology etc.

References

1. Agrawal, R., Faloutsos, C., Swami, A.: Efficient Similarity Search in Sequence Databases. In: Lomet, D.B. (ed.) FODO 1993. LNCS, vol. 730, pp. 69–84. Springer, Heidelberg (1993)
2. Agrawal, R., Lin, K.-I., Sawhney, H.S., Shim, K.: Fast similarity search in the presence of noise, scaling, and translation in time-series databases. In: 21st International Conference on Very Large Databases, pp. 490–501. Morgan Kaufmann, San Francisco (1995)
3. Batyrshin, I., Herrera-Avelar, R., Sheremetov, L., Panova, A.: Moving approximation transform and local trend associations in time series data bases. In: Batyrshin, I., Kacprzyk, J., Sheremetov, L., Zadeh, L. (eds.) Perception-based Data Mining and Decision Making in Economics and Finance. SCI, vol. 36, pp. 55–83. Springer Physica Verlag (2007)

4. Batyrshin, I., Sheremetov, L., Velasco-Hernandez, J.X.: On axiomatic definition of time series shape association measures. In: Workshop on Operations Research and Data Mining, ORADM 2012, Cancun, pp. 117–127 (2012)
5. Batyrshin, I., Bulgakov, I., Hernandez, A.-L., Huitron, C., Chi, M., Raimundo, A., Cosultchi, A.: VMD-Petro®: Visualization and data mining tool for oilfields. In: Workshop on Operations Research and Data Mining, ORADM 2012, Cancun, pp. 140–148 (2012)
6. Batyrshin, I.Z., Sheremetov, L.B.: Perception-based approach to time series data mining. Applied Soft Computing 8, 1211–1221 (2008)
7. Chatfield, C.: The Analysis of Time Series: An Introduction. Chapman and Hall (1984)
8. Das, G., Gunopulos, D.: Time series similarity and indexing. In: Handbook on Data Mining, pp. 279–304. Lawrence Erlbaum Associates (2003)
9. Fu, T.-C.: A review on time series data mining. Engineering Applications of Artificial Intelligence 24, 164–181 (2011)
10. Goldin, D.Q., Kanellakis, P.C.: On Similarity Queries for Time-Series Data: Constraint Specification and Implementation. In: Montanari, U., Rossi, F. (eds.) CP 1995. LNCS, vol. 976, pp. 137–153. Springer, Heidelberg (1995)
11. Kacprzyk, J., Wilbik, A., Zadrozny, S.: Linguistic summarization of trends: a fuzzy logic based approach. In: 11th Int. Conf. Information Processing and Management of Uncertainty in Knowledge-based Systems, IPMU 2006, pp. 2166–2172 (2006)
12. Last, M., Kandel, A., Bunke, H.: Data Mining in Time Series Databases, Machine Perception and Artificial Intelligence, vol. 57. World Scientific (2004)
13. Liao, T.W.: Clustering of time series data – A survey. Pattern Recognition 38, 1857–1874 (2005)
14. Möller-Levet, C.S., Klawonn, F., Cho, K.-H., Wolkenhauer, O.: Fuzzy Clustering of Short Time-Series and Unevenly Distributed Sampling Points. In: Berthold, M., Lenz, H.-J., Bradley, E., Kruse, R., Borgelt, C. (eds.) IDA 2003. LNCS, vol. 2810, pp. 330–340. Springer, Heidelberg (2003)
15. Rafiei, D., Mendelzon, A.O.: Querying time series data based on similarity. IEEE Transactions on Knowledge and Data Engineering 12, 675–693 (2000)